Technical Report

NetApp Seismic Processing Solution Guide

Robert Lai, NetApp
June 2012 | TR-3959

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

1.1 NetApp Seismic Processing Solution Introduction

Overview

High-performance computing (HPC) applications for oil and gas exploration enable the processing of extremely large amounts of very rich data, such as prestack seismic data, which can involve simulation times over long periods. Prestack data can be hundreds of terabytes and sometimes even petabytes in size before the final seismic imaging applications are completed. Today's highly complex seismic data acquisition systems employ new techniques, such as wide-azimuth towed streamer (WATS) and coil shooting for off-shore data acquisition, resulting in full azimuth (FAZ) seismic data, which is stored for later processing. All of the data-generation techniques can create hundreds of gigabytes per hour. To effectively manage, use, and quickly access this data, companies require a proven, high-performance, enterprise-reliable data storage solution. NetApp has invested in innovative storage technologies, the right partnerships, and a wealth of industry expertise to be the tier-1 solution provider to the oil and gas industry.

Among numerous requirements is the need for a single file system that is capable of accessing the entire set of seismic data. A scalable file system is required to handle the size as well as the raw storage performance for processing.

In order for seismic data processing to yield reliable results, the solution must have the following key features:

- Well-defined high-performance storage portfolio at attractive prices
- High-density storage enclosures scaling to 1PB and beyond
- Highly scalable file system performance
- Reliability, availability, and serviceability

The NetApp® Seismic Processing Solution (SPS) meets all of these requirements by offering a reference configuration that is suitable for the seismic processing industry. Reference configurations scale to meet the most demanding application requirements by using common storage building blocks to expand and meet any combination of performance and/or capacity needs.

NetApp Benefits

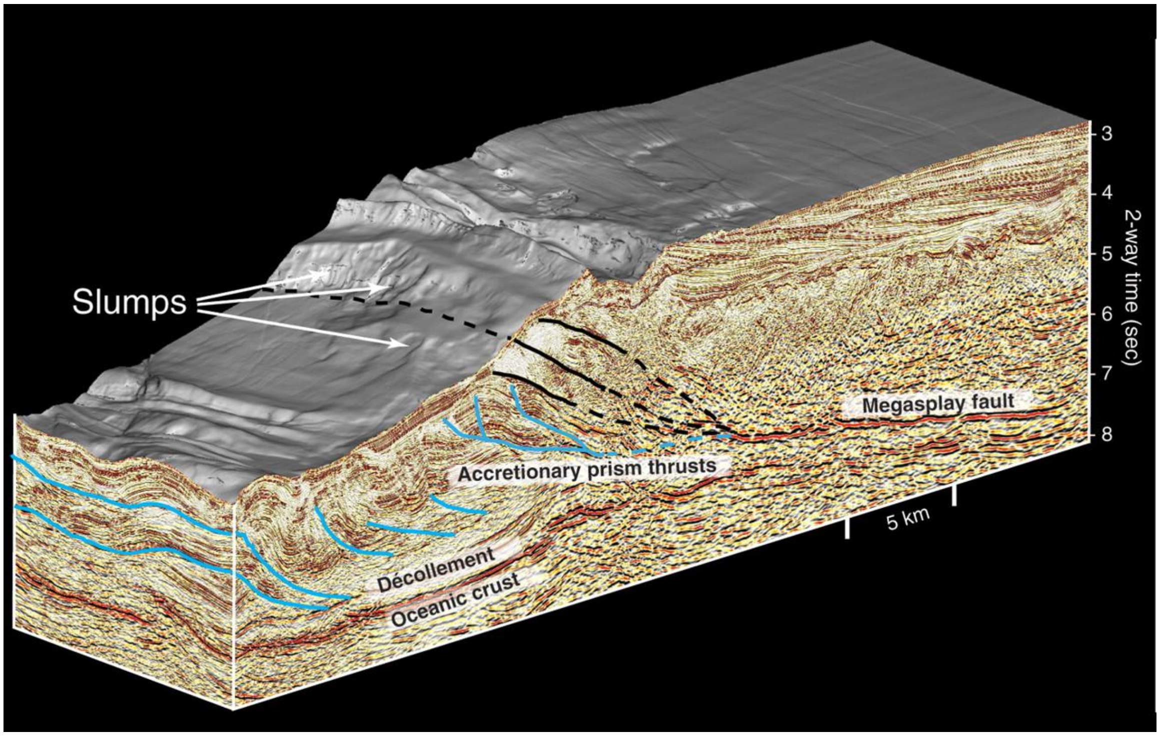
The oil and gas industry requires data storage for a complex and diverse group of professionals. Geologists, geophysicists, engineers, landmen, accountants, lawyers, and field personnel require fast and consistent access to data. Additionally, each of these professions has working categories, such as reservoir engineers, drilling engineers, operations engineers, and environmental engineers, all of whom need seamless access to myriad data formats. The E5460 and E5424 (E5400) storage systems offer reliable and innovative storage solutions for this dynamic industry.

The E5400 storage system offers the following features:

- Bandwidth to easily handle the rigors of heavy computational workloads and bandwidth-sensitive environments
- Reliable systems for fast and confident deployments; the advanced thermal, power, and failover features minimize downtime
- Standard components that offer redundancy, reliably delivering critical data
- Manageability offered by the Quantum® StorNext® File System (SNFS) integration to deliver streamlined management for large datasets
- Scalability to accommodate growing data streams and access requirements

Collecting and processing large datasets obtained by a variety of both on-shore and off-shore seismic acquisition systems is essential to the upstream exploration industry. The NetApp E5400 storage system's mature architecture, integrated with Quantum SNFS, offers an economical and reliable HPC storage environment suitable for the high throughput required for seismic processing. Figure 1 illustrates the complexity of seismic data.

Figure 1) Real-time seismic data analysis.



Key features of the NetApp E-Series storage solutions include:

- SPS handles the real-time seismic data throughput of today's seismic imaging applications, from Kirchhoff prestack depth migration (PSDM) all the way to tilted transverse isotropy (TTI) reverse time migration (RTM) HPC codes.
- SPS is able to deliver the prestack data and velocity files to HPC clusters made up of hundreds or thousands of compute cores.
- SPS has high-density storage that accommodates the many travel-time cubes or even the large n-dimensional JavaSeis sparse file format.
- Systems are tested with seismic imaging applications that result in higher confidence.
- Standard components offer redundancy, reliably delivering critical data.
- Integration with SNFS delivers streamlined management for large datasets.
- Scalability accommodates growing data streams, longer retention policies, and access requirements.
- SPS is designed to meet the service demand for metadata and large file sets used in applications such as 3D and 4D seismic processing and other related workloads.
- SPS has a variety of supported drive types for high-performance and high-capacity dataset needs.

Architecture and Components

The SPS consists of E-Series storage and the Quantum StorNext shared file system. Optional components include SAN hosts, a SAN fabric, LAN clients, and archive storage. NetApp Professional Services and SupportEdge are required solution components.

A generic SPS is composed of the components listed in Table 1.

Table 1) Seismic Processing Solution component list.

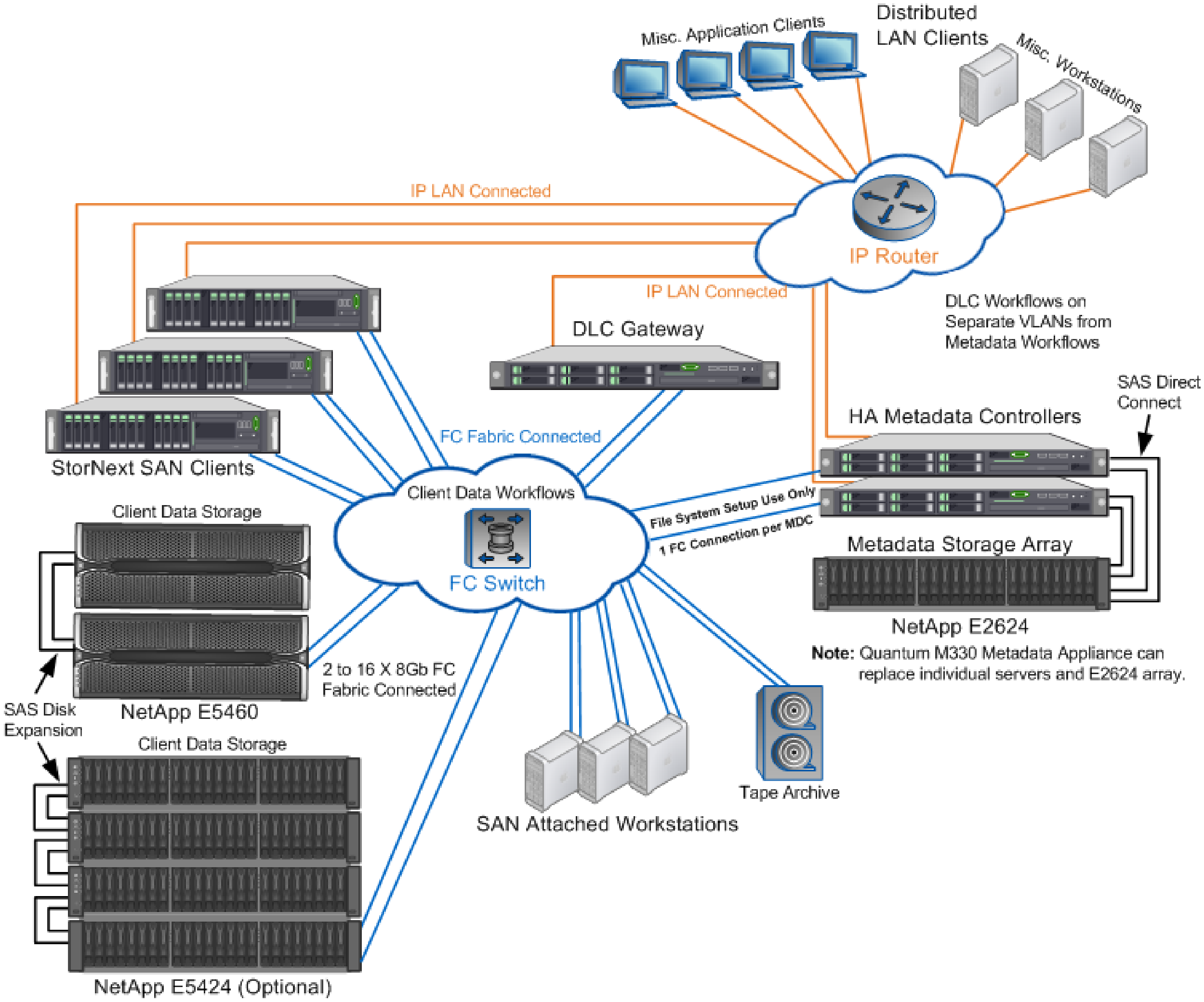
SPS Components	Description	Hardware	Supplier
Application/user storage	One or more 60-drive or 24-drive storage chassis.	NetApp E5460, E5424	NetApp
Metadata controller (MDC)	SPS uses the Quantum SNFS. The MDC server makes metadata available to clients. Optionally, Quantum offers the M330 appliance that bundles two MDC servers plus metadata storage.	A Linux® server (two servers for high-availability [HA] mode)	Third-party server supplier or Quantum
Metadata storage	Storage for housing the file system metadata and journal data. Usually configured on storage that is separate from the application/user storage. Quantum offers the M330 appliance that bundles two MDC servers plus metadata storage.	NetApp E2624	NetApp, Quantum, other
SPS SAN hosts	The StorNext driver must be installed on the host to allow it to mount the file system.	A Linux server or a Windows® server	Integrator or customer
StorNext distributed LAN server (DLS)	A DLS server acts as a gateway for LAN clients that require access to the file system.	A Linux server or a Windows server	Integrator or customer
SPS LAN clients	Typically, Linux servers with a StorNext driver installed. The servers are likely to run over a 10GbE network.	Linux, Windows, or Mac® OS X computers	Integrator or customer
SAN fabric	Fibre Channel (FC) switch and cabling to provide input/output (I/O) access to E-Series storage.		Integrator or customer
Gigabit Ethernet network	Dedicated network for Quantum StorNext MDC, hosts, and LAN clients.		Integrator or customer

SPS Components	Description	Hardware	Supplier
Second- and third-tier storage	Quantum StorNext offers optional methods for archiving data to disk or tape.	Additional storage and/or tape library archive	Integrator or customer

For more information, refer to the datasheet for the NetApp E5400 Storage System.

Figure 2 depicts a simplified overview of SPS as part of a media workflow.

Figure 2) Typical Seismic Processing Solution architecture components.



E-Series Architecture

The NetApp SPS consists of the E5460 and/or E5424 E-Series storage systems and the Quantum SNFS and data management functions. These storage systems feature dual E5400 RAID controllers in either the DE6600 4U 60-drive shelf or the DE5600 2U 24-drive shelf. Each shelf can be populated with NL-SAS, SAS, or SSD drives.

The E5460 and E5424 are fifth-generation storage arrays that include patented mechanical engineering and provide dense, scalable, and highly reliable bandwidth and capacity. The disk controller firmware supports an optimal mix of high-bandwidth large-block streaming and small-block random I/O.

A base E5460 or E5424 may be expanded with the addition of one or several corresponding DE6600 or DE5600 expansion enclosures. The DE6600 and DE5600 are disk expansion enclosures or shelves that hold disks but no RAID controllers. These are daisy-chained to the E5460 or E5424 and provide expansion storage behind the RAID controllers in the base unit. The SPS can be architected to independently scale capacity and bandwidth to best meet customer requirements.

The E2624 consists of a DE5600 expansion enclosure but uses the E2600 RAID controller instead of the E5400. This storage system may be used to store the Quantum SNFS metadata and journal data.

Table 2 provides guidelines for E-Series expansion options for SPS.

Table 2) E-Series disk expansion shelf guidelines.

Category	E5460	E5424	E2624
Form factor	4U/60 drives	2U/24 drives	2U/24 drives
Maximum disk drives	360	192	24
Controller shelf	1	1	1
Maximum expansion shelves	5	7	0
Total number of disk shelves	6	8	1

Quantum StorNext File System Architecture

The Quantum SNFS provides a large single data container, performance aggregation, and heterogeneous shared-data access. StorNext provides SAN, IP, and NAS client (NAS requires setting up a gateway server) connectivity. Client bandwidth is enhanced through third-party FC host bus adapters (HBAs), StorNext-optimized LAN protocol for Microsoft® Windows and Linux clients, and StorNext-optimized Apple® Filing Protocol (AFP) server technology for Apple Mac clients.

The Quantum SNFS provides block-based access to the same underlying file structure housed on a single external storage system or striped across multiple storage systems.

The two major components of a Quantum SNFS environment are the MDC and a client that is installed on each host requesting access to the shared file system. Communication between the MDC and clients occurs through a segregated Ethernet network intended for use only by the StorNext MDC and file system clients. This private network does not have to provide large amounts of bandwidth, but NetApp highly recommends a very low-latency private Ethernet connection between the client nodes and the MDC.

The MDC is responsible for arbitrating overall access to the shared file system, as well as for configuration and management. It can be configured with either a single server or dual servers for HA through additional redundancy. The MDC software is supplied as a distribution based on either Linux or Microsoft Windows.

The StorNext client software is installed on each server or workstation designated for access to the file system. Mount and unmount operations are largely controlled by the client software and its communication with the MDC.

For more information, refer to the StorNext File System page on the Quantum Web site.

Conclusion

The NetApp SPS is designed and optimized for the most demanding seismic data processing workloads. The preconfigured, pretested solution is designed to support the high bandwidth required to capture and share large imaging datasets across multiple applications and sites. By enabling faster data exploitation, organizations have the information needed to accelerate exploration and production operations.

- Big bandwidth support delivers up to 4.4GB/sec bandwidth in a single 4U rack unit.
- Modular design allows growth with minimal components, eliminating the need to overconfigure.
- High density supports 1.8PB in each industry-standard 40U rack.
- Cost-effective expansion allows scaling of bandwidth and capacity independently within the same container. Start small and expand with 2U or 4U increments as demand changes.

2 Solution Overview

2.1 E-Series Quantum StorNext Solutions Sizing Considerations

Overview

Sizing is a critical component of architecting an E-Series Quantum StorNext solution, and it begins with determining capacity and throughput goals. In order to properly size a Quantum StorNext application, it is important to gather as much information and as many requirements as possible. These are categorized as follows:

- Application environment, user population, applications, and workflows
- Application data storage capacity required by users and their applications
- Overall throughput performance required by the application environment and estimates for the associated number of concurrent sequential I/O streams
- Growth factors for future data storage and performance
- Media format selected by the users and their applications

Workflow Requirements

There are many varied application types, use cases, and resulting I/O patterns in Quantum StorNext environments. For sizing purposes, some simplifying assumptions must be made about the characteristics of parallel I/O that are relevant for most Quantum StorNext environments. In general, the highest throughput performance is required when clients of the file system are executing parallel sequential, large-request-size reads and writes with the Quantum SNFS. These sequential file reads and writes are referred to as streams.

To the E-Series Quantum StorNext solutions storage elements, this application I/O typically looks like many distributed, concurrent read and write streams spread across the storage systems. The E-Series Quantum StorNext solutions have been measured under similar test conditions to characterize the throughput performance across a wide range of concurrent sequential I/O streams for reads and writes of various I/O sizes.

Capacity Requirements

For either existing applications usage or new application deployment, it is necessary to determine the total amount of storage capacity required across the file system. This amount is usually specified in terabytes. Also note the number of users and the total number of files in use.

Performance Requirements

Performance is usually specified as the total storage throughput required at peak periods. This is measured in existing application environments and in new environments as the intended gigabytes per second (GB/sec) in aggregate for all applications.

Also important for performance is the number of concurrent I/O streams that are executed to achieve the overall throughput. It is necessary to use both total throughput and the number of streams in calculations for sizing. Also of note are the percentages relative to read and write for the total throughput requirements. Currently, sizing is calculated only for the total throughput required for the aggregate of read and write streams.

Metadata performance has also been characterized for such operations as:

- Number of file creations and deletions per second
- Directory creations and deletions per second
- Number of stats per second (file or directory information, such as that provided in Linux)

This performance is relatively independent of E-Series storage array performance and is primarily dictated by the choice of hardware for the MDC and by the StorNext metadata architecture.

Calculation of Storage Sizing Requirements for Quantum StorNext

There are two sizing calculations for Quantum StorNext user data and metadata storage to support the Quantum SNFS. The user data storage requirements are specified by customers for their specific environments. It may be necessary to calculate the metadata storage required if current storage requirements are not available. This calculation requires the total number of files per directory, the total file system size, and the file system block (FSB) size for the Quantum SNFS.

Quantum StorNext Metadata

For Quantum StorNext metadata storage, NetApp recommends the E2624, which is configured to provide maximum metadata performance and to satisfy capacity requirements. A single E2624 shelf with 24 600GB 10K RPM hard-disk drives (HDDs) provides the capacity to support petabyte-sized file systems (depending on average file size) and meets storage performance requirements such that metadata performance is limited by the MDC server performance, not by the E-Series storage performance.

Calculating Metadata Storage Requirements

The capacity required for the Quantum SNFS metadata depends on the combination of:

- Desired overall file system size
- Defined FSB size
- Approximate number of files per directory

Use the information provided in Table 3 to help calculate the required metadata capacity.

Table 3) Metadata capacity requirements.

Average Number of Files per Directory	File System Size: Less than 10TB	File System Size: 10TB or Larger
Less than 10	FSB: 16KB Metadata: 32GB per 1M files	FSB: 64KB Metadata: 128GB per 1M files
10–100	FSB: 16KB Metadata: 8GB per 1M files	FSB: 64KB Metadata: 32GB per 1M files

Average Number of Files per Directory	File System Size: Less than 10TB	File System Size: 10TB or Larger
100–1,000	FSB: 64KB Metadata: 8GB per 1M files	FSB: 64KB Metadata: 8GB per 1M files
More than 1,000	FSB: 64KB Metadata: 4GB per 1M files	FSB: 64KB Metadata: 4GB per 1M files

In addition to this metadata calculation, journal storage requirements must be added. From a disk space perspective, the journal portion of a Quantum SNFS requires relatively little storage space. This space is often no larger than 500MB.

NetApp recommends the E2624 with 24 600GB HDDs as the standard storage configuration for StorNext metadata storage. The recommended best practice for RAID configurations for metadata is to use RAID 10 with multiple drives, a 128KB segment size, and two hot spare drives.

Multiple metadata and journal logical unit numbers (LUNs) can be configured for use by multiple SNFS MDCs on the E2624 metadata storage array. Each file system requires metadata and journal storage, both on the same volume. With the recommended configuration, 22 drives are available for RAID 10 volume groups, and they may be used to support different file systems. Metadata performance testing was done using a 2+2 configuration as an example.

For the recommended storage configuration, Table 4 gives the total maximum capacity of the E2624 that can be used in multiple RAID 10 volume groups for multiple file systems.

Table 4) E2624 metadata capacity example using RAID 10.

E-Series	Controller Shelf	Drives	Drive Size	Raw	Formatted (RAID 10)
E2600	E2624	24	600GB	14.4TB	6.0TB

Given the available capacity of 6TB, this metadata storage can support extremely large SNFSs supporting many millions of files.

Estimating Metadata Performance

With RAID 10 volumes using 600GB SAS drives and the recommended configuration settings, the metadata performance is limited by the choice of server for the MDC and is not limited by the storage performance of the E2624 array. For sizing purposes, capacity is the only required parameter since performance is not a factor.

Metadata performance was tested with a dual-socket Westmere-class server with 48GB of memory as the MDC and an E2624 with 24 2.5" SAS 600GB 10K RPM HDDs in the recommended configuration for metadata storage performance of RAID 10 using 2+2 drives with a 128KB segment size. Table 5 lists the expected input/output operations per second (IOPS) for performance elements associated with metadata operations in this configuration.

Table 5) Metadata performance statistics.

Operation	Performance (IOPS)
Directory creation	116
File creation	1550

Metadata performance numbers vary depending on the type of server used, but performance is not limited by the E2624 metadata storage.

Quantum StorNext User Data Storage

Sizing E-Series storage for Quantum StorNext user data begins with the dual considerations of capacity and performance requirements. For the purpose of user data, NetApp offers two platform choices:

- E5460 with 3.5" 7.2K RPM HDDs
- E5424 with 2.5" 10K RPM HDDs

Each platform provides specific performance, capacity, and scale-out capabilities.

This guidance assumes that the customer has already selected a preferred basic drive technology, either 3.5" SAS 7.2K RPM or 2.5" SAS 10K RPM HDDs. The trade-offs between these two technologies include form factor, capacity per disk, performance, and other features. In general, for environments that require large capacities and large-format sequential throughput, NetApp recommends selecting the 3.5" 7.2K RPM drives. For higher stream counts and when performance is a critical requirement, NetApp recommends selecting the 2.5" 10K RPM drives. Sizing rules are provided for both drive types.

Sizing for performance is done first, and it determines the number of controller shelves needed to meet the requirement. For each controller shelf, LUNs are configured and used by the file system in StorNext stripe groups, which define the set of LUNs over which files are striped by StorNext. Capacity is then calculated for the resulting controller shelves required for performance. If this capacity does not meet the overall capacity requirement, extra expansion shelves are added to scale out to the required capacity.

This guidance also assumes that there is linear performance scaling as StorNext stripe groups are added to the file system. Although this is a reasonable assumption for StorNext in general, there are many reasons that a collection of client servers might not achieve 100% linear performance scaling with additional stripe groups. There are issues of choice of the SAN clients, network design and performance, distribution of files and workloads across those stripe groups, and varying approaches to determining precisely the way applications perform I/O and interact with the shared file system. These issues are beyond the scope of this document but are the subject of future work planned for this solution.

Sizing for User Data Storage Performance

The storage performance sizing requirement has two components:

- Total storage I/O bandwidth, or throughput, required across the entire SNFS, usually stated in GB/sec
- Maximum number of concurrent I/O streams from all clients across the entire file system to achieve the total aggregate throughput

Testing was developed to measure the NetApp E-Series Quantum StorNext solutions' storage performance and to create metrics to use for sizing calculations that are representative of customer use cases and environments. Most Quantum StorNext customer environments present storage workloads composed of concurrent file read and write streams that, in aggregate, define the required system throughput. Testing produced a synthetic workload with ranges of concurrent I/O streams and I/O sizes to simulate these customer environments.

From this testing, two operating regions were identified and metrics were developed to describe throughput versus I/O count when sizing storage systems that will operate predictably in the regions. In the first region, testing revealed that maximum saturated performance for the E-Series controllers was achieved for small to moderate numbers of concurrent streams. This is the preferred operating region for environments that require the highest performance. Application workloads are constrained to operate below the maximum number of supported concurrent streams for this region to achieve this performance metric.

As the number of concurrent streams is increased, the streams intermix and present a more random I/O workload to the storage. In region 1, the low-to-moderate stream count region, the StorNext SAN client and E-Series controllers are able to process the intermixed concurrent I/O streams and still achieve the maximum bandwidth saturated throughput of the controllers. Above the maximum stream count for region

1, the I/O randomness results in drive-limited performance, defined in region 2. This is the performance metric for workloads that require large numbers of concurrent I/O streams.

Sizing for performance is determined by the total aggregate throughput required and the total number of concurrent I/O streams at that throughput. Either region 1 (moderate stream counts) or region 2 (large stream counts) will produce the optimal number of controller shelves, depending on the total required concurrent streams.

The following sizing calculations are provided for these two regions:

- Inputs:
  - Array type – E5460 or E5424
  - Aggregate throughput required in GB/sec – MaxThroughput
  - Total concurrent I/O streams at aggregate throughput – MaxStreams
- Outputs:
  - Required controller shelves of either type E5460 or type E5424
  - Minimum drive and shelf requirements to achieve the required performance
- Calculations:
  - Total controller shelves required = Lesser of [region 1 controller shelves or region 2 controller shelves]

Refer to Table 6 to determine maximum streams and performance values per array model and configuration for region 1 and region 2 calculations:

- Region 1 controller shelves = Greater of [MaxThroughput / (region 1 controller shelf performance for the array type) or MaxStreams / (region 1 max streams)]
- Region 2 controller shelves = Greater of [MaxThroughput / (region 2 controller shelf performance for the array type) or MaxStreams / (region 2 max streams)]

Table 6 provides the metrics for region 1 (moderate number of streams) and region 2 (high number of streams) performance sizing.

Table 6) E-Series performance region metrics.

Array Model	Number of Drives	RAID Format	Region 1: Maximum Streams	Region 1: Controller Shelf Performance (GB/sec)	Region 2: Maximum Streams	Region 2: Controller Shelf Performance (GB/sec)
E5460	30	8+2 RAID 6	300	1.4	3,000	0.6
E5460	60 to 360	8+2 RAID 6	600	2.8	6,000	1.2
E5424	24	7+1 RAID 5, 6+2 RAID 6	300	1.4	3,000	0.6
E5424	48 to 192	7+1 RAID 5, 6+2 RAID 6	600	2.8	6,000	1.2

Performance can be sensitive to the record size used for the reads and writes. These sizing metrics were tested for record sizes between 128KB and 4096KB.

Sizing for User Data Storage Capacity

When sizing for performance is complete, the resulting capacity can be calculated and compared against the overall capacity requirement. If the resulting capacity is insufficient, additional expansion shelves with drives can be added based on the extra capacity necessary to meet the requirement.

For the E5460, the recommended best practice and configuration used in this document is RAID 6 (8+2) with a 128KB segment size for a total RAID stripe size of 1MB. Settings for best practices and best performance for the file system configuration and client configuration are given in the Quantum StorNext File System Tuning Guide.

For the E5424, NetApp recommends two configurations, based on customer preference for RAID 5 or RAID 6. For RAID 5, the best practice and configuration used in this document is RAID 5 (7+1) with a 128KB segment size for a total RAID stripe size of 896KB. For RAID 6, the best practice and configuration used in this document is RAID 6 (6+2) with a 128KB segment size for a total RAID stripe size of 768KB. Settings for best practices and best performance for the file system configuration and client configuration are given in the Quantum StorNext File System Tuning Guide.

The E5460 supports either three or six RAID 6 (8+2) volume groups using either a half-populated (30-drive) shelf or a fully populated (60-drive) shelf. This is the same configuration used for expansion with the DE6600 expansion shelf.

The E5424 supports either three RAID 5 (7+1) or three RAID 6 (6+2) volume groups (for a total of 24 drives) per shelf. These are the same configurations used for expansion with the DE5600 expansion shelf.

Table 7 lists the capacity sizes for various HDD and drive shelf options.

Table 7) Capacity calculations for E5460 and E5424 configurations.

E-Series Controller	Controller Shelf	Expansion Shelf	Drives	Drive Sizes RAID	Raw (TB)	Formatted (TB)
E5400	E5460 (half populated)		30	2/3TB 8+2 RAID 6	60/90	43.7/65.5
ESM		DE6600 (half populated)	30	2/3TB 8+2 RAID 6	60/90	43.7/65.5
E5400	E5460		60	2/3TB 8+2 RAID 6	120/180	87.3/130.9
ESM		DE6600	60	2/3TB 8+2 RAID 6	120/180	87.3/130.9
E5400	E5424		24	600GB/900GB 7+1 RAID 5	14.4/21.5	11.3/17.6
ESM		DE5600	24	600GB/900GB 7+1 RAID 5	14.4/21.5	11.3/17.6
E5400	E5424		24	600GB/900GB 6+2 RAID 6	14.4/21.5	9.7/15.1
ESM		DE5600	24	600GB/900GB 6+2 RAID 6	14.4/21.5	9.7/15.1

When configuring expansion shelves (DE6600, DE5600), it is important to observe the maximum drive counts per controller shelf system in order not to exceed allowable drive limits for the respective controllers. Table 8 lists the maximum drive counts and shelves.

Table 8) Maximum drive counts and shelves.

E-Series Controller	Controller Shelf	Expansion Shelf	Maximum Number of Drives	Maximum Number of Expansion Shelves Supported
E5400	E5460		60 (360 total per array maximum)	5 (6 total shelves per array maximum)
ESM		DE6600	300	5
E5400	E5424		24 (192 total per array maximum)	7 (8 total shelves per array maximum)
ESM		DE5600	168	7
E2600	E2624		24 (192 total per array maximum)	7 (8 total shelves per array maximum)
ESM		DE5600	168	7

It is possible to mix expansion shelves (DE6600, DE5600) with the rule that the maximum number of drives cannot exceed the controller shelf limit. If expansion shelves or drive count maximums are exceeded, it is necessary to configure additional controller shelf systems to meet the desired capacity requirements.

Note: Mixing shelf models requires dedicated proof-of-concept testing that is not part of the standard test process for this solution.

Once the sizing for performance is completed and the number of controller shelves with drives is calculated, the capacity for this performance storage is calculated based on the information in Table 7. If the overall capacity required exceeds the capacity achieved through performance sizing, additional drive shelves and drives can be added to achieve the overall capacity. The total number of expansion shelves that can be attached to a controller shelf is limited, as Table 8 indicates.

Each time a RAID 6 (8+2) volume group is added, it is also added to a StorNext stripe group for use by the file system. Additional stripe groups added through expansion drive shelves should be allocated evenly across each array in order to maintain symmetrical file system performance and to achieve symmetrical file system capacity across the entire file system. In the end, the best practice is for each array to have an identical number, type, and capacity of LUNs, unless there are multiple file systems with different stripe group requirements and a requirement to aggregate storage as densely as possible. However, NetApp recommends keeping the design as symmetrical as possible for best performance and performance sizing.

If the overall desired capacity cannot be achieved after the maximum number of expansion shelves behind each controller shelf is reached, additional controller shelves with symmetrical expansion shelves must be added to reach the overall required capacity.

Sizing Example

In this example, a customer needs an SNFS configuration that supports 8.8GB/sec sequential I/O throughput for 300 concurrent streams, with a total user storage capacity of 1PB. Throughput performance is the primary consideration for the environment, so planning to operate in region 1 as defined in Table 6 (that is, 2.8GB/sec) is required. The customer is using RAID 6 (8+2) and wants to use the E5460, with 60 3TB drives installed in each DE6600 shelf. In this example, the NetApp Systems

Engineer or Professional Services Engineer must use the following calculations to determine the number of shelves needed to satisfy the throughput and capacity requirements.

Capacity Calculation

The customer uses the following capacity calculation to determine the total number of disk shelves required, including controller shelves:

1. Convert the required capacity from petabytes to terabytes using the conversion factor of 1024TB = 1PB (that is, 1PB x 1024TB / 1PB = 1024TB).
2. Determine the formatted capacity for a 60-disk shelf with 60 x 3TB drives installed (130.9TB per shelf of formatted capacity, as shown in the Formatted (TB) column in Table 7).
3. Divide the total terabytes of required capacity by the value of formatted capacity per shelf configuration:
   - (1024TB / X shelves) = (130.9TB / 1 shelf)
4. Solve for X using the full equation:
   - [(1PB x 1024TB / 1PB) / X shelves] / (130.9TB / 1 shelf)

The result is X = 1024TB / 130.9TB = 7.8. Therefore, in this example, the customer needs 8 shelves.

Throughput Calculation

The customer uses the following throughput calculation to determine the number of controller shelves required for 8.8GB/sec of overall throughput:

1. Determine the maximum sequential throughput for an E5460 shelf (2.8GB/sec for up to 600 streams, as shown in Table 6):
   - (300 streams required / 600 streams maximum for the E5460 in region 1) = .5 controller shelves
2. Determine the maximum throughput for this configuration (2.8GB/sec per shelf for region 1, as shown in Table 6):
   - 8.8GB/sec / 2.8GB/sec = 3.14 shelves (round up to 4 shelves)
3. To satisfy the performance requirement, the customer must use the larger number from these two throughput calculations, which is 4 controller shelves.

Expansion Calculation

The customer uses the following expansion calculation to determine the required number of expansion shelves:

1. Determine the capacity for the required number of controller shelves:
   - 130.9TB x 4 shelves = 523.6TB
2. Determine the additional expansion capacity required:
   - 1024TB – 523.6TB = 500.4TB
3. Determine the number of expansion shelves required for the additional capacity:
   - 500.4TB / 130.9TB per DE6600 = 3.8 shelves (round up to 4 DE6600 expansion shelves)

Based on these calculations and on the best-practice recommendation to keep the design symmetrical, the expansion capacity needed is 1 DE6600 expansion shelf with 60 3TB drives behind each of the 4 controller shelves, for a total of 8 shelves (4 E5460 controller shelves plus 4 DE6600 expansion shelves). With a total of 480 3TB drives, the configuration in this example provides an overall performance of 11.2GB/sec and 1047.2TB of capacity and 1047.2TB of formatted disk capacity.

Scale-Out Capacity and Performance Sizing

Sizing Quantum SNFS storage capacity and performance involves calculating the necessary number of controller shelves and expansion shelves for the users' data storage. This guidance assumes linear performance scaling as stripe groups are added. Likewise, linear capacity scaling is achieved as controller shelves and expansion shelves are added. This straightforward scale-out methodology is a key attribute of the NetApp E-Series product line with the Quantum SNFS.

Summary

The NetApp E-Series E5460, E5424, and E2624 storage systems provide the necessary capacity, throughput, IOPS, and response times to meet performance requirements for demanding Quantum SNFS environments.

2.2 E-Series Quantum StorNext Solutions Performance Considerations

Performance considerations are an essential element when planning to architect and implement E-Series Quantum StorNext solutions. Understanding the typical workloads of the solution environments in which E-Series arrays provide the critical storage infrastructure underpinning high-performance file systems is essential to determine how to tune the storage, host, and file system settings to maximize performance for those workloads.

Overview

Building a test methodology that is representative of real-world workloads is key to characterizing performance and enabling optimization. Although there are many different applications in the Quantum StorNext solutions environment, these applications have some commonality in what they require from the shared file system environment. Video, media, and seismic processing applications generally must perform parallel reads and writes to and from the file system and into and out of the memories of the client nodes.

Although strategies differ for how these applications store information on the file system, a generalized approach for achieving optimal parallel I/O performance is for each process to write or read its information using its own dedicated single file. This minimizes contention between nodes and processes for reading and writing files in parallel and allows the distribution of I/O streams across the entire storage system for maximum throughput.

Since each of these processes typically opens a file and performs a complete read or write, the resulting I/O stream is 100% sequential read or write and the Quantum SNFS client aggregates requests and tries to produce aligned, large, well-formed I/O for best throughput.

Performance Characterization for Quantum StorNext

The planning for solutions based on Quantum StorNext must account for and characterize two aspects of E-Series storage performance:

- Performance for user data I/O to volumes within stripe groups
- Metadata operations performance to the metadata and journal volume stripe group

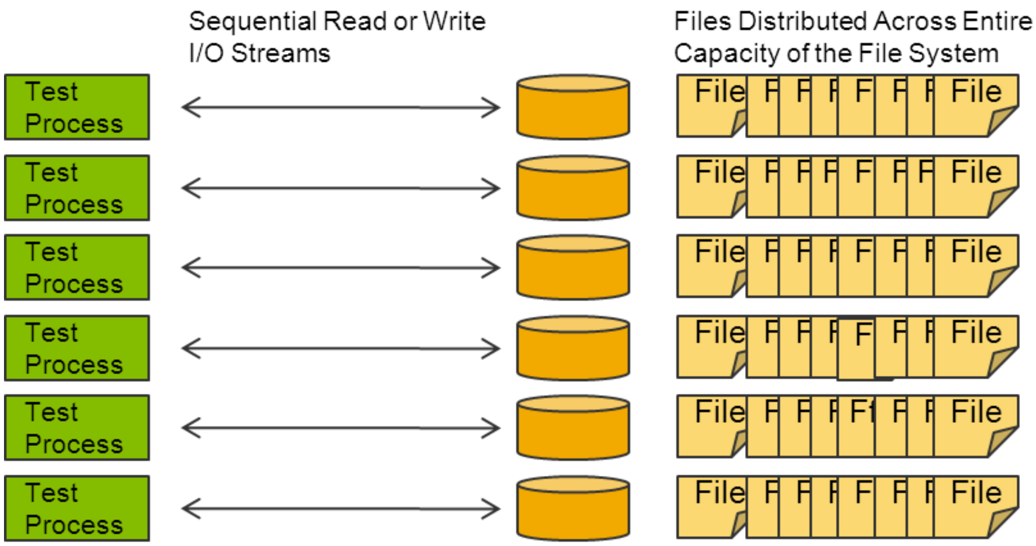
The methodology for testing user data stripe group performance characterizes shared file system throughput using volumes on E5460 or E5424 storage. The requirement for user data stripe group performance is for optimal large I/O sequential throughput from multiple concurrent streams.

The methodology for testing metadata performance characterizes the IOPS behavior of the metadata server using volumes on the E2624 as metadata storage. Metadata storage operations are typically random small-block I/O, and they require optimal IOPS performance from the metadata storage configuration.

Quantum StorNext User Data Storage Performance

The test methodology used for characterizing and optimizing solution performance emulates the process-per-file strategy. A multiserver, file-based I/O benchmarking tool is used to generate a range of file read and write streams to files configured across the entire capacity of the file system. Test tools, such as vdbench and iozone, are used to provide the ability to format files on the file system and distribute file read and write processes across a number of test client nodes, synchronize the I/O activities across these nodes, and coalesce results into summary data files. Figure 3 illustrates the methodology.

Figure 3) E-Series Quantum SNFS performance test methodology.



The methodology for testing file read and write data path performance across the file system includes the following tasks:

- Format thousands of files of sufficient size to use the entire file system capacity and therefore to span all sectors of all drives in the test file system.
- Run a series of automated test scripts on a number of client nodes that measure throughput with a number of processes, each reading or writing sequential streams to individual files across the file system.
- Run these tests for ranges from one to thousands of total streams for each client connected to the E-Series storage array.
- At each test case for a number of streams, run various I/O sizes and run from 100% reads to 100% writes, as well as combinations in between.

During testing, these automated tests were run across a range of storage configurations and across a range of numbers of stripe groups and volumes per stripe group. The tests were run for single and multiple E5460 or E5424 controller shelves and with a range of drive counts and expansion shelves. The aggregate throughput results for both reads and writes were plotted against the number of streams to determine the performance behavior for a range of file-per-process I/O streams that Quantum StorNext solutions applications might provide.

Figure 4 shows a qualitative view of this behavior for 100% read streams, 100% write streams, mixtures of read and write streams, and various I/O sizes.

Figure 4) E-Series performance zones.

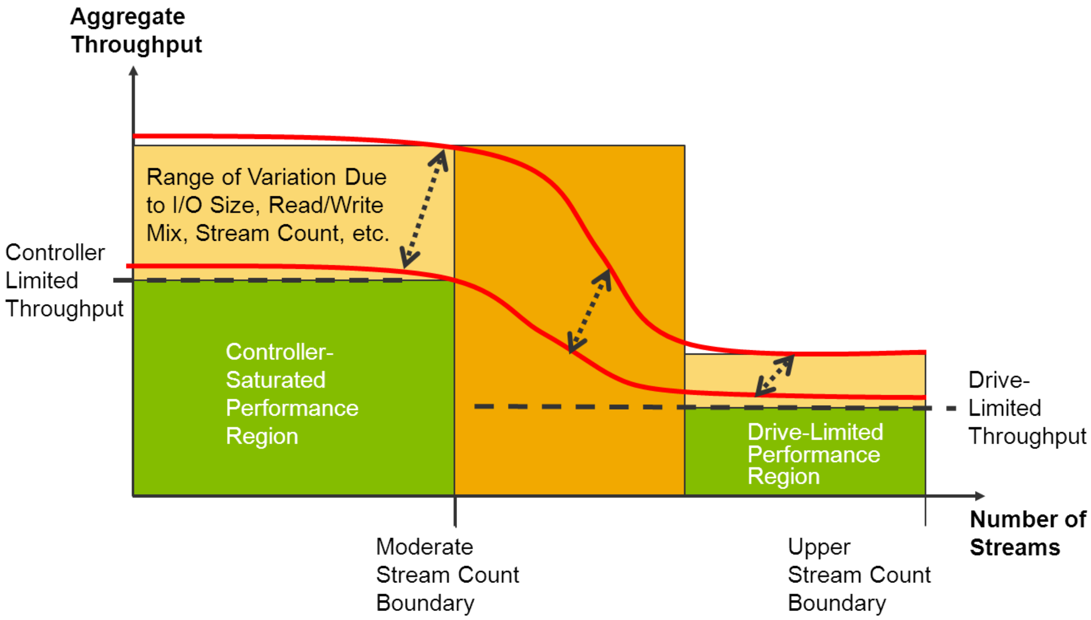


Figure 4 shows aggregate throughput versus stream count for a single E-Series controller shelf, with bands of values for various I/O sizes and combinations of read and write streams. This figure also shows performance from the perspective of applications on the clients making file read and write system calls on the client server nodes. The regions indicate the overall response of the SNFS clients, the SAN network, and the user data storage mounted by those StorNext SAN clients. This provides an integrated view of system throughput, which is similar to the throughput of an actual application doing concurrent sequential file reads and writes.

The first green region demonstrates the value and efficiency of the NetApp E-Series Quantum StorNext–based solutions by showing that applications can yield controller-saturated performance for low to moderate stream counts for a variety of I/O sizes and mixtures of read and write streams. This saturation of the controllers can be achieved because buffers in the StorNext SAN client and caches in the E-Series RAID controllers are able to aggregate and reorder I/O sufficiently to provide optimal, well-formed I/O requests to the HDDs in the E-Series systems. This well-formed I/O to the disks means that the HDDs are used in streaming mode and can keep pace with the throughput of the controllers. Optimal I/O formatting is important in sizing because it simplifies potentially complex customer architectures by providing a flat performance level that can be achieved for each controller shelf over a range of conditions. Total aggregate performance can be achieved by adding controller shelves using a linear throughput scale-out approach enabled by the SNFS.

The green controller-limited throughput region is determined by selecting the test result for the condition that creates the minimum saturated controller limited throughput, adjusted to account for test and manufacturing variations. This produces a high-confidence sizing number that works for a wide range of conditions and provides the value used in the sizing algorithm.

In Figure 4, the first yellow region above the controller-saturated green region shows performance results that can be achieved through optimizations such as selecting specific mixes of read and write percentages or specific I/O sizes. Customer designs may plan for sizing in this region if specific I/O conditions are met, but there is increased risk that performance could vary and might not meet requirements. For designs that might operate in the yellow region, proof-of-concept testing of specific environments is required to demonstrate that performance objectives can be achieved as stream counts increase past the moderate stream count boundary, since the combination of larger numbers of concurrent I/O streams spread widely across the address space makes it increasingly difficult to keep the

HDDs in fully streaming mode. As stream counts continue to increase to the upper stream count boundary, the aggregate controller shelf throughput becomes drive limited and represents the large I/O random performance of the HDDs in the E-Series system. For sizing purposes, NetApp advises using large-stream-count sizing rules in the darker yellow region where throughput is making the transition from controller-saturated to drive-limited performance.

In the second green region in Figure 4, the aggregate throughput is determined by selecting the test result for the condition that produces the minimum drive-limited throughput, adjusted to account for test and manufacturing variations. Sizing rules in this region set the aggregate throughput to the drive-limited value for stream counts from the moderate stream count limit to the upper stream count limit.

The second yellow region corresponds to variations above the drive-limited throughput caused by mixes of I/O sizes and reads versus writes. Again, if sizing optimization requires operating in this region, a proof-of-concept test is required in order to have confidence in the expected performance.

E5460 Performance Test

The E5460 system was tested for the Quantum StorNext user data stripe groups using both 30- and 60-drive configurations attached to StorNext SAN clients. The configuration for distributed LAN clients (DLCs) and distributed LAN servers (DLSs) was not tested and is not within the scope of this document.

Test Environment

The E5460 was tested with 60 drives configured in six 10-drive RAID 6 (8+2) volume groups. These six volume groups were used to provide six total volumes (utilizing all of the capacity of each volume group) to the Quantum SNFS as user data storage. Testing was also done with a half-populated controller shelf with 30 drives installed and configured for three RAID 6 (8+2) volume groups and volumes for the file system user data storage. The hardware segment size was always set at 128K, and all caching (controller read and write caching with mirroring) was enabled.

Results and Analysis

The minimum aggregate read and write throughput measured for the E5460 with six RAID 6 (8+2) LUNs used, with 100 or fewer streams, was 2.8GB/sec; 100% reads were measured at 5GB/sec and 100% writes were measured at 2.8GB/sec. These measurements correspond to the controller-saturated green performance region on the performance chart. For low stream counts, the yellow controller-saturated performance region shows that the 2.8GB/sec may be exceeded under circumstances in which the read and write ratio approaches 100% reads.

For stream counts from 100 to 1,000, the minimum aggregate bandwidth was measured to be no less than 1.2GB/sec.

A half-shelf configuration (30 3.5" 3TB drives) was also tested. For 50 or fewer steams, the aggregate bandwidth was measured at no less than 1.4GB/sec, and for 50 to 500 streams this value was no less than 0.6GB/sec.

Adding expansion shelves and more than 60 drives did not increase the controller-saturated performance results. In general, adding drives does have a positive impact on the moderate stream count value (increasing it) and does improve the drive-limited performance since there are more drive spindles and IOPS to use in a random workload environment. However, it is more effective to achieve higher performance by adding controller shelves than by adding expansion shelves. Therefore, NetApp does not recommend adding expansion shelves to achieve performance gains when throughput performance is the main objective, since adding controller shelves is far more effective. Although NetApp recommends expansion shelves and additional drives for adding capacity, in testing for sizing purposes no additional performance gains were observed.

Conclusion

Each E5460 controller shelf (60 3TB drives) provides a minimum of 2.8GB/sec of aggregate read and write bandwidth for all combinations of read and write percentages at stream counts from 6 to 100 streams per stripe group and provides a minimum of 1.2GB/sec for stream counts from 100 to 1,000 streams per stripe group. The workload stream must be capable of driving the data in order for these results to remain valid.

E5424 Performance Test

The E5424 array may also be used for StorNext user data storage using 2.5" SAS drive technology. Optionally, customers can use this array if their capacity needs do not require 3.5" drives and if they want the higher drive performance, smaller footprint, and lower power consumption of this system.

Test Environment

For the Quantum StorNext performance metrics, the E5424 was configured with 24 600GB SAS drives running at 10K RPM. These drives were configured in three sets of eight-drive RAID 6 (6+2) volume groups and, separately, three sets of eight-drive RAID 5 (7+1) volume groups. No hot spares were assigned. Three volumes were provided to the Quantum SNFS for user data storage from the controller shelf, and an additional three volumes were provided to the Quantum SNFS through an expansion shelf. The configuration was tested to reach saturated controller performance.

Results and Analysis

The minimum aggregate read and write throughput for the E5424 (three RAID 5 [7+1] LUNs or RAID 6 [6+2] LUNs) was measured at 1.4GB/sec for 50 or fewer streams. Based on test results, 100% reads are specified at 1.4GB/sec and 100% writes are specified at 1.8GB/sec.

For stream counts from 50 to 500, with three LUNs in the E5424 controller shelf, the aggregate read and write throughput was no less than 0.6GB/sec.

Doubling the number of drives with the use of an E5424 + DE5600 (48 drives total) configured with six (RAID 5 [7+1] or RAID 6 [6+2]) LUNs essentially doubled the performance metrics. The minimum aggregate read and write throughput for the E5424 + DE5600 was 2.8GB/sec for 100 or fewer streams. Based on test results, 100% reads are specified at 3.5GB/sec and 100% writes are specified at 2.8GB/sec.

For stream counts from 100 to 1,000 with six LUNs, the aggregate read and write throughput was no less than 1.2GB/sec.

Adding expansion shelves beyond the controller shelf (48 drives total) did not increase the controller-saturated performance results. In general, adding more than 48 drives does have some impact on the moderate stream count value (increasing it) and does improve the drive-limited performance since there are more drive spindles and IOPS to use in a random workload environment. However, it is more effective to achieve higher performance by adding controller shelves than by adding expansion shelves. Therefore, NetApp does not recommend adding expansion shelves to achieve performance gains when throughput performance is the main objective, since adding controller shelves is far more effective. Although NetApp recommends expansion shelves and additional drives for adding capacity, in testing for sizing purposes no additional performance gains were observed.

Conclusion

Each 24-drive E5424 controller shelf provides 1.4GB/sec of aggregate bandwidth for all combinations of read and write streams, for stream counts from 1 to 50. Doubling the drive count (that is, providing a second drive shelf) to use 48 drives provides 2.8GB/sec of aggregate bandwidth for all combinations of read and write streams, for stream counts from 1 to 100.

Summary of Performance and Sizing Guidelines

Table 9 summarizes the performance and sizing guidelines for the E5460 and E5424, based on the number of drives, the number of concurrent streams, and the resulting performance. It includes numbers for region 1 (a moderate number of streams) and region 2 (a high number of streams).

Table 9) E-Series performance region metrics.

Array Model	Number of Drives	RAID Format	Region 1: Maximum Streams	Region 1: Controller Shelf Performance (GB/sec)	Region 2: Maximum Streams	Region 2: Controller Shelf Performance (GB/sec)
E5460	30	8+2 RAID 6	300	1.4	3,000	0.6
E5460	60 to 360	8+2 RAID 6	600	2.8	6,000	1.2
E5424	24	7+1 RAID 5, 6+2 RAID 6	300	1.4	3,000	0.6
E5424	48 to 192	7+1 RAID 5, 6+2 RAID 6	600	2.8	6,000	1.2

Quantum StorNext Metadata Performance

Metadata storage is provided by the E2624 and consists of a choice of 600GB or 900GB SAS drives that run at 10K RPM. These drives are configured into RAID 10 LUNs, which are presented to the Quantum SNFS.

Metadata performance is primarily a function of the performance of the metadata server and—when configured correctly—is not limited by the E2624 storage. The performance of metadata operations is characterized here to confirm that the storage is not the limiting factor and to provide some guidance for performance expectations using a typical metadata server.

Test Environment

For the Quantum StorNext metadata performance metrics, the E2624 was configured with 24 600GB drives that run at 10K RPM. For testing, LUNs were created and configurations were tested as shown in Table 10.

Table 10) Metadata storage configuration tested for performance characterization.

RAID Level	Disk Configuration	Segment Size
RAID 1	1+1	32K
RAID 1	1+1	64K
RAID 1	1+1	128K
RAID 10	2+2	32K
RAID 10	2+2	64K
RAID 10	2+2	128K
RAID 10	4+4	32K
RAID 10	4+4	64K

RAID Level	Disk Configuration	Segment Size
RAID 10	4+4	128K

Each LUN was used in a separate test as the metadata and journal stripe group for a single file system. There was very slight performance improvement with the 1x1 LUN with the 128k segment size, but overall there were extremely small differences using the various configurations.

Results and Analysis

Dmake and fmake tests provided by Quantum were used to measure the rate at which files and directories could be created per second. Table 11 lists the metadata performance statistics.

Table 11) Metadata performance statistics.

Operation	Performance (IOPS)
Directory creation	116
File creation	1550

Conclusion

StorNext best practices specify that the metadata/journal stripe group should be placed on a RAID 1 or a RAID 10 LUN that uses the full capacity of the disks associated with the LUN. The E-Series best metadata performance was observed at a 128k segment size. Therefore, NetApp recommends using a 128k segment size as the initial configuration setting. No significant performance improvement over RAID 1 was seen with RAID 10 volumes.

Scale-Out Performance Considerations

For StorNext user data storage, the E5460 and the E5424 storage systems can be expanded by using expansion shelves to meet capacity and throughput requirements. Capacity requirements are met through the use of multiple shelves with the appropriate drives for the E5460 or the E5424. The number of controller shelves depends on the throughput and stream count requirements.

Summary

The E-Series E5460, E5424, and E2624 storage systems provide the necessary capacity, throughput, IOPS, and response times to meet performance requirements for demanding Quantum SNFS environments.

2.3 E-Series Solutions Hardware Packaging

Table 12 lists the part numbers associated with all E-Series solutions. These are the part numbers that are included in the EzChoice Quote Tool.

Table 12) E-Series part numbers.

Category	Part Number	Product Description
System enclosures	DE6600-SYS-ENCL-R6	Enclosure, 4U-60, DE6600, empty, 2PS
	DE5600-SYS-ENCL-R6	Enclosure, 2U-24, DE5600, empty, 2PS
	DE1600-SYS-ENCL-R6	Enclosure, 2U-12, DE1600, empty, 2PS

Category	Part Number	Product Description
Expansion enclosures	E-X5680A-QS-R6	Enclosure, 4U-60, DE6600, empty, 2PS, QS
	E-X5681A-QS-R6	Enclosure, 2U-24, DE5600, empty, 2PS, QS
	E-X5682A-QS-R6	Enclosure, 2U-12, DE1600, empty, 2PS, QS
ESM controller	E-X30030A-R6	ESM controller, SBB-2
5400 controllers	E5400A-12GB-R6	E5400A, 12GB controller
	E5400A-6GB-R6	E5400A, 6GB controller
2600 controller	E2600A-2GB-R6	E2600A, 2GB controller
Host interface cards (HICs)	X-52708-00-R6	HIC, E5400, 40GB, IB, 2-port
	X-48855-00-R6	HIC, FC, 4-port, 8Gb, E5400
	X-52709-00-R6	HIC, E2600, 1GB iSCSI, 4-port
	X-52710-00-R6	HIC, E2600, 10GB iSCSI, 2-port
	X-52194-00-R6	HIC, E2600, FC, 4-port, 8Gb
	X-52195-00-R6	HIC, E2600, SAS, 2-port, 6Gb
Racks	X-M102061-R6	40U rack, empty, L6-30, domestic
	X-M102062-R6	40U rack, empty, IEC309, international
Software (point of sale)	SW-5400-FDE-SKM-P	SW, full-disk encryption (FDE) security key management, 5400, -P
	SW-5400-SNAPSHOT-P	SW, Snapshot, 5400, -P
	SW-5400-VOL-COPY-P	SW, volume copy, 5400, -P
	SW-5400-REM-MIRR-P	SW, remote mirroring, 5400, -P

Category	Part Number	Product Description
	SW-2600-FDE-SKM-P	SW, FDE sec key management, 2600, -P
	SW-2600-SNAPSHOT-P	SW, Snapshot, 2600, -P
	SW-2600-VOL-COPY-P	SW, volume copy, 2600, -P
	SW-2600-REM-MIRR-P	SW, remote mirroring, 2600, -P
Software (add-on)	SW-5400-FDE-SKM	SW, FDE sec key management, 5400
	SW-5400-SNAPSHOT	SW, Snapshot, 5400
	SW-5400-VOL-COPY	SW, volume copy, 5400
	SW-5400-REM-MIRR	SW, remote mirroring, 5400
	SW-2600-FDE-SKM	SW, FDE sec key management, 2600
	SW-2600-SNAPSHOT	SW, Snapshot, 2600
	SW-2600-VOL-COPY	SW, volume copy, 2600
	SW-2600-REM-MIRR	SW, remote mirroring, 2600
Disk drives (packs)	E-X4021A-10-R6	Disk drives, 10x3TB, 7.2k, DE6600
	E-X4023A-10-R6	Disk drives, 10x2TB, 7.2k, DE6600
	E-X4022A-12-R6	Disk drives, 12x3TB, 7.2k, DE1600
	E-X4024A-12-R6	Disk drives, 12x2TB, 7.2k, DE1600
	E-X4027A-12-R6	Disk drives, 12x600GB, 3.5", 15k, DE1600
	E-X4025A-12-R6	Disk drives, 12x900GB, 2.5", 10k, DE5600
	E-X4026A-12-R6	Disk drives, 12x600GB, 2.5", 10k, DE5600
Single disk drives	E-X4021A-R6	Disk drive, 3TB, 7.2k, DE6600, QS
	E-X4023A-R6	Disk drive, 2TB, 7.2k, DE6600, QS

Category	Part Number	Product Description
	E-X4022A-R6	Disk drive, 3TB, 7.2k, DE1600, QS
	E-X4024A-R6	Disk drive, 2TB, 7.2k, DE1600, QS
	E-X4027A-R6	Disk drive, 600GB, 3.5", 15k, DE1600, QS
	E-X4025A-R6	Disk drive, 900GB, 2.5", 10k, DE5600, QS
	E-X4026A-R6	Disk drive, 600GB, 2.5", 10k, DE5600, QS
Expansion 10 packs	E-X4021A-10-QS-R6	Disk drives, 10x3TB, 7.2k, DE6600, QS
	E-X4023A-10-QS-R6	Disk drives,10x2TB, 7.2k, DE6600, QS
Expansion 12 packs	E-X4022A-12-QS-R6	Disk drives, 12x3TB, 7.2k, DE1600, QS
	E-X4024A-12-QS-R6	Disk drives, 12x2TB, 7.2k, DE1600, QS
	E-X4027A-12-QS-R6	Disk drives, 12x600GB, 3.5", 15k, DE1600, QS
	E-X4025A-12-QS-R6	Disk drives, 12x900GB, 2.5", 10k, DE5600, QS
	E-X4026A-12-QS-R6	Disk drives, 12x600GB, 2.5", 10k, DE5600, QS
Spares/field replaceable units	E-X4028A-R6	Solid-state drive, 800GB, 2.5", DE6600, QS
Disk drives single (field-replaceable unit only)	E-X4029A-R6	Solid-state drive, 200GB, 2.5", DE6600, QS
	E-X4030A-R6	Solid-state drive, 800GB, 2.5", DE5600, QS
	E-X4031A-R6	Solid-state drive, 200GB, 2.5", DE5600, QS
Miscellaneous hardware	X-48788-00-R6	Controller, E5400, 12GB, FC, no battery, SMID161
	X-24238-00-R6	Rail kit, DE1600, adjustable, 23.5"–32.5"
	X-41198-00-R6	Rail kit, DE6600, adjustable, 29.5"–35.75"
	X-48601-00-R6	Rail kit, DE6600, adjustable, 600–785mm

Category	Part Number	Product Description
	X-48870-00-R6	PSU, 725W, AC, DE1600
	X-48564-00-R6	PSU, 1755W, AC, DE6600
	X-46381-00-R6	Battery, E2600
	X-48619-00-R6	Battery, E5400
	X-48565-00-R6	FAN, DE6600
	X-48566-00-R6	Drawer, 12-drive, DE6600
	X-48567-00-R6	Bezel, front panel, DE6600
	X-24936-00-R6	Cable, mini-SAS, 2m, R6
	X-37953-00-R6	SFP, 8Gb, FC, E-Series
Power cords	X-50613-00-R6	Power cord, in-cabinet, 2m, C14-C19, 250V, DE6600
	X-52197-00-R6	Power cord, in-cabinet, 2m, C14-C13, E-Series
	X-33106-00-R6	Power cord, North America, 220V, E-Series
	X-33107-00-R6	Power cord, North America, 110V, E-Series
	X-33108-00-R6	Power cord, Europe, E-Series
	X-33109-00-R6	Power cord, Switzerland, E-Series
	X-33110-00-R6	Power cord, Italy, E-Series
	X-33111-00-R6	Power cord, UK and Ireland, E-Series
	X-33112-00-R6	Power cord, Denmark, E-Series
	X-33113-00-R6	Power cord, India, E-Series
	X-33115-00-R6	Power cord, Australia-New Zealand, E-Series
	X-33116-00-R6	Power cord, Israel, E-Series
	X-33117-00-R6	Power cord, China, E-Series
	X-41592-00-R6	Power cord, Taiwan, E-Series

3 Management of E-Series

3.1 E-Series SANtricity ES Management Client Out of Band

SANtricity® ES software is the GUI used to manage E-Series storage arrays. The application is based on the Java® framework and can be installed on Windows or Linux operating systems.

Install the management application on a management node that does not participate in the data delivery workload. The installation package supports both 32- and 64-bit machines. The installation procedure verifies that the application is being installed on the correct OS version.

Overview

The SANtricity ES Management Client in the out-of-band configuration enables storage administrators to perform the following tasks:

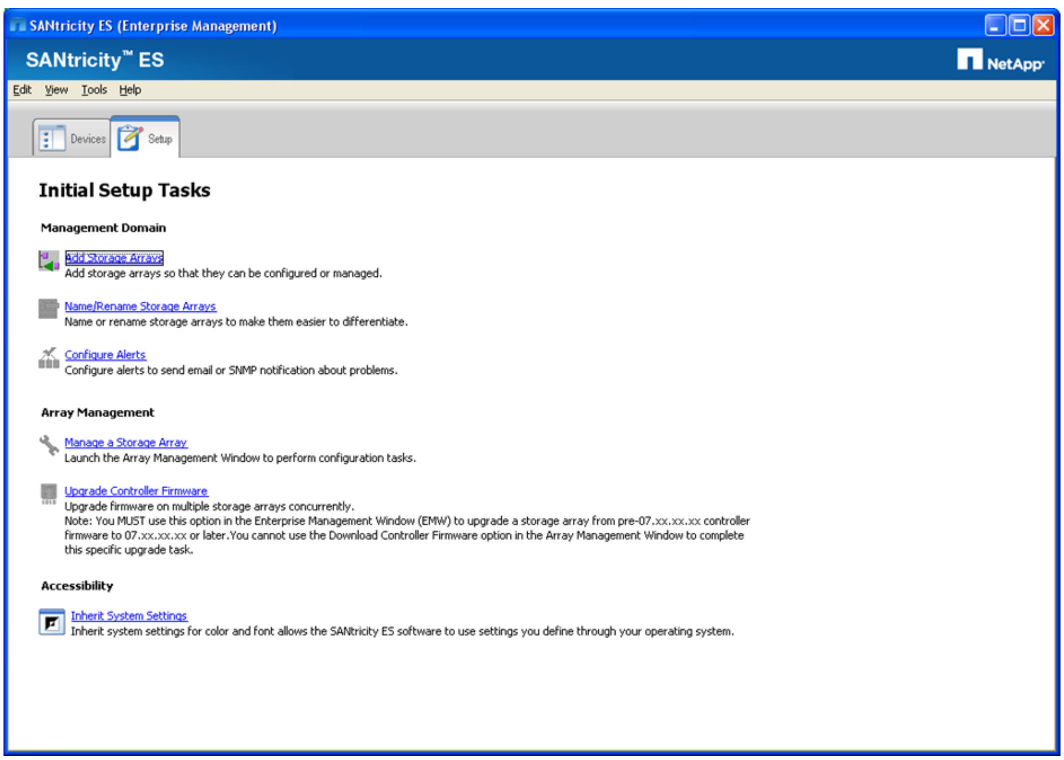
- Commission new storage devices.
- Set up network connections.
- Provision storage and hosts.
- Perform various maintenance activities to manage storage on E-Series storage arrays.

When the SANtricity ES Management Client is installed on a desktop OS, the following limitations apply:

- Simultaneous user sessions are limited to eight sessions.
- Real-time system monitoring is not enabled.
- Desktop systems cannot run the host agent and send I/O to the E-Series storage array.

When the SANtricity ES Management Client is installed on a compute platform running a server OS, the full functionality is available; however, the number of simultaneous sessions is limited. Figure 5 shows the SANtricity ES EMW landing page.

Figure 5) SANtricity ES Management Client EMW.



Guidelines

Apply the following guidelines to use the SANtricity ES Management Client in the out-of-band configuration:

- Use out-of-band management when the storage administrator segregates management I/O from production I/O, and during the initial commissioning operations that occur before hosts are connected to the array.
- The management server must access the storage array through an IP connection (Dynamic Host Configuration Protocol [DHCP] or static) to the Ethernet management ports on the controller modules.
- The out-of-band management client cannot run the real-time monitoring feature.
- Install the management application on a management node that does not participate in the data delivery workload.

Note: SANtricity ES is supported on management servers with FC or SAS connectivity to the storage array; however, it must have connectivity to each array to be properly managed. Refer to the SANtricity ES 10.80 online help documentation for more information about SANtricity ES in the in-band configuration.

4 Physical Infrastructure for E-Series

4.1 E-Series E5400 Hardware

Overview

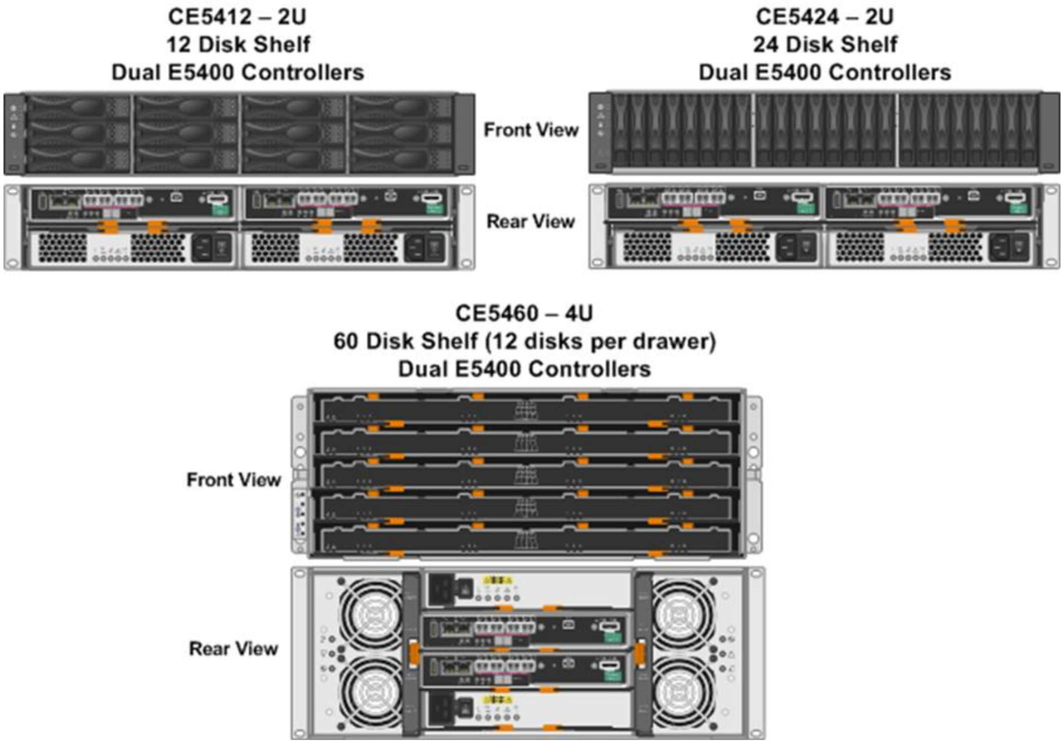
High-bandwidth applications and HPC platforms require high-performance, reliable, and scalable storage systems. The E5400-based storage system meets these requirements by supporting:

- Four 8Gb FC host interfaces per controller standard
- Multiple optional HICs, one per controller:
  - Four-port 8Gb FC
  - Two-port 40Gb IB
- 384 total disk drives per storage array
- Multiple RAID levels (0, 1, 10, 3, 5, and 6)
- A range of drive speeds and capacities
- Data assurance (T10-PI data integrity checking)
- Media parity check and correction capability
- Extensive event logging
- Recovery Guru onboard system diagnostics and recovery capability
- Hardware redundancy
- 6GB cache memory per controller (12GB optional) to maximize read/write performance
- NVSRAM and onboard USB drive to preserve the system configuration during power outages

As shown in Figure 6, the E5400 controller is available in three shelf packages (E5460, E5424, and E5412), each supporting dual-controller canisters, power supplies, and fan units for hardware redundancy. The shelves are sized to support 60 disks, 24 disks, or 12 disks, respectively. Multiple disk expansion shelves (DE6600, DE5600, and DE1600) can be connected to the controller shelf to add additional storage capacity. For additional details, refer to the NetApp E5400 Storage System datasheet.

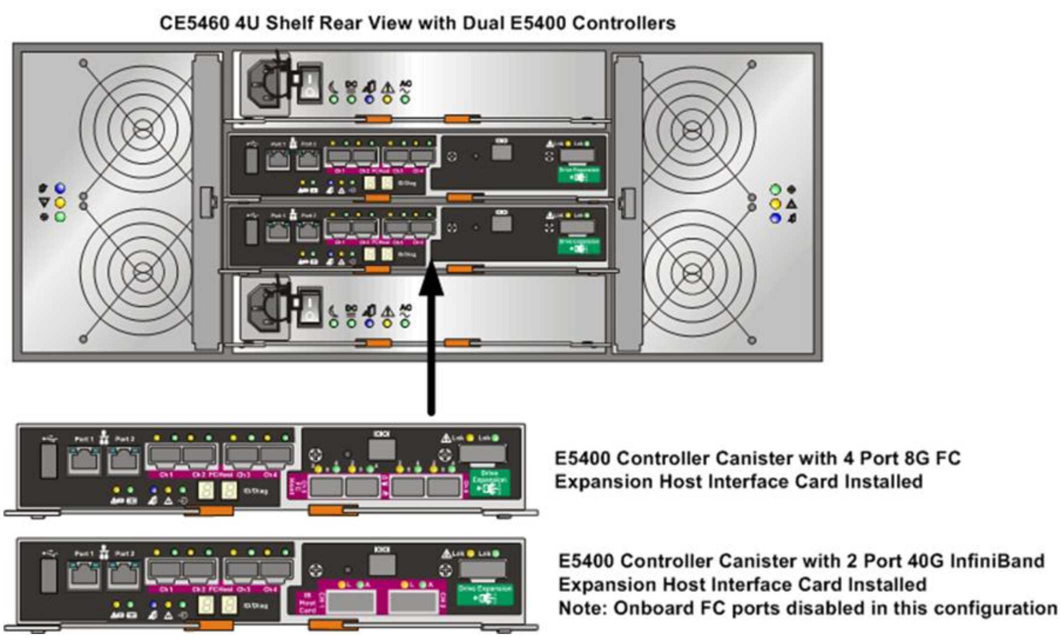
Note: The DE6600 60-disk E5400-based arrays should not exceed 6 total shelves counting the controller shelf, and the DE5600 and DE1600 shelf configurations should not exceed 8 total shelves. Empty slots in any attached disk shelf are counted as drives when calculating the total drive count on an array.

Figure 6) E5400 shelf options.



By default, the E5400 controller canister has four 8Gb onboard FC ports for host-side communication channels, but it also supports channel expansion through add-on modules that add either four FC ports or two IB ports (onboard FC ports are disabled when the IB module is installed). Figure 7 shows the E5460 4U shelf with the available channel adapter modules.

Figure 7) E5460 controller shelf with optional host-side expansion ports.

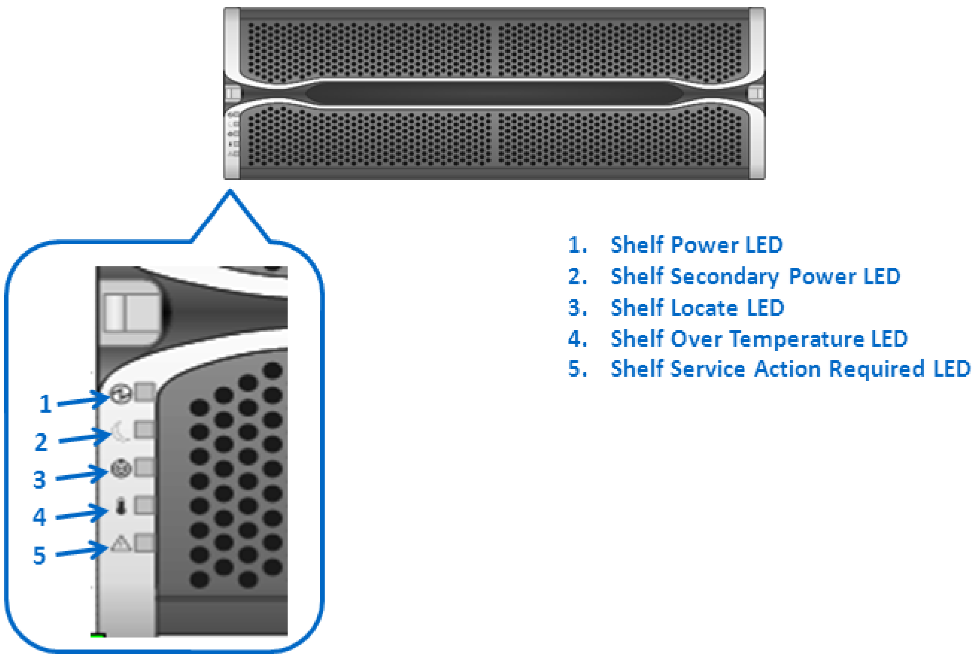


LED Status Indicators

Controller Drive Shelf LED Status Indicators

The E5400 controller shelf enclosure has several light-emitting diodes (LEDs) that indicate the overall status of the array, as shown in Figure 8.

Figure 8) Controller drive shelf status LEDs.



The shelf status LED layout is the same for all three packaging options (DE6600, DE5600, and DE1600). Table 13 lists the meanings of all the indicators.

Table 13) Controller drive shelf LED status definitions.

LED Name	Color	LED On	LED Off
Controller Drive Shelf Power	Green	Power is present.	Normal status.
Controller Drive Shelf Secondary Power	Green	Battery is fully charged. LED blinks when battery is charging.	Controller canister is operating without battery, or existing battery has failed.
Controller Drive Shelf Locate	White	Identifies controller-drive tray when SANtricity ES Locate feature is activated.	Normal status.
Controller Drive Shelf Over Temperature	Amber	The temperature of the controller-drive tray has reached unsafe level.	Normal status.
Controller Drive Shelf Service Action Required	Amber	A component within the controller-drive tray requires attention.	Normal status.

Controller Base Features LED Status Indicators

The E5400 controller has several onboard LED status indicators, as shown in Figure 9. Most of the LEDs are lit when a fault condition exists; however, the battery charging and the cache-active LEDs are lit when the battery is fully charged and the cache is active. The seven-segment LEDs provide status codes for both normal operation and fault conditions, and the dot in the first seven-segment LED is the controller heartbeat indicator.

Figure 9) E5400 controller status indicator LEDs.

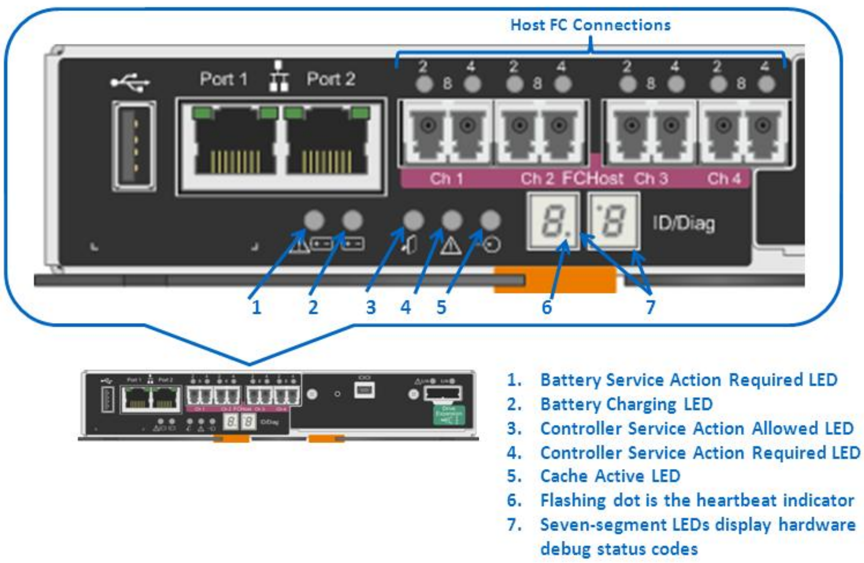


Table 14 provides additional controller status LED definitions.

Table 14) Controller base features LED status definitions.

LED Name	Color	LED On	LED Off
Battery Service Action Required LED	Amber	Battery in controller canister has failed.	Normal status.
Battery Charging LED	Green	Battery is fully charged. LED blinks when battery is charging.	Controller canister is operating without battery, or existing battery has failed.
Controller Service Action Allowed LED	Blue	Controller canister can be removed safely from controller drive tray.	Controller canister cannot be removed safely from controller drive tray.
Controller Service Action Required LED	Amber	Some fault exists within controller canister.	Normal status.
Cache Active LED	Green	Cache is active. After AC power failure, this LED blinks while cache offload is in process.	Cache is inactive, or controller canister has been removed from controller-drive tray.
Dot in Lower-Right Corner of First Seven-Segment LED	Yellow (not amber)	Dot flashing indicates controller heartbeat is active.	Dot not lit indicates controller heartbeat is not active (that is, controller is not in service).

LED Name	Color	LED On	LED Off
Two Seven-Segment LEDs	Yellow (not amber)	If controller status code = 99, then controller is in service. If controller status code does not = 99, then fault condition exists. Contact Technical Support for further assistance.	Controller is not powered on.

Note: The Battery Service Action Required LED indicates that the battery timer has expired or the battery has failed the automatic battery test. This condition can seriously impact system write performance because the write cache feature is automatically disabled when the battery is not functioning normally.

Host-Side Ports LED Status Indicators

The host-side connection ports provide status LEDs to indicate the connection status for each link between the storage array and various host-side hardware devices, as shown in Figure 9. Table 15 and Table 16 provide the definitions for each LED.

Table 15) Ethernet management port status indicator definitions.

LED Name	Color	LED On	LED Off
Ethernet Management Port Link Rate LED (top left corner of Mgmt. port RJ-45 connectors)	Green	There is a 100BASE-T rate.	There is a 10BASE-T rate.
Ethernet Management Port Connectors Link Active LED (top right corner of Mgmt. port RJ-45 connectors)	Green	Link is up (LED blinks when there is activity).	Link is not active.

Table 16) Host-side FC ports status indicator definitions.

FC Port LEDs (Link Active and Data Rate)	Color	LED On
Upper Left = Off, Upper Right = Off	Green	Link not active
Upper Left = On, Upper Right = Off	Green	Link active, data rate = 2Gb/sec
Upper Left = Off, Upper Right = On	Green	Link active, data rate = 4Gb/sec
Upper Left = On, Upper Right = On	Green	Link active, data rate = 8Gb/sec

Drive-Side SAS Expansion Port

The E5400 controller canister is equipped with a 4-lane 6Gb/s SAS expansion port used to connect additional disk shelves to the E5400 controller shelf. Figure 10 shows a close-up of the SAS expansion port LEDs.

Figure 10) E5400 drive expansion port status indicator LEDs.

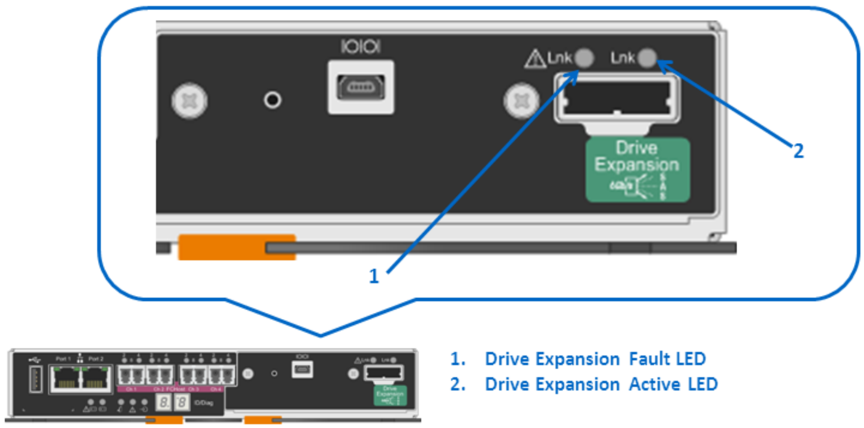


Table 17 provides the definitions for each drive-side LED.

Table 17) Drive-side SAS ports status indicator definitions.

LED Name	Color	LED On	LED Off
Drive Expansion Link Fault	Amber	At least one of the four PHYs in Out port is working, but another PHY cannot establish same link to Expansion Out connector.	Normal status.
Drive Expansion Link Active	Green	At least one of four PHYs in Out port is working, and link exists to device connected to Expansion Out connector.	Link error has occurred.

For additional details on the E5400 controller and related hardware, refer to the NetApp E-Series Storage Systems CE5400 Controller-Drive Tray Installation Guide.

Guidelines

Consider the following guidelines when implementing the E5400 storage system:

- Determine the level of performance required by the compute platforms to support the given applications.
- Determine the amount of storage capacity required (include the number of disks required for hot spares).
- Choose the disk types based on performance and capacity requirements.
- Determine the power and network connectivity requirements.
- Plan RAID levels to achieve the level of reliability and read/write performance required.
- Determine which hosts will be connected to the storage system, and plan the configuration of the storage system ports to maximize throughput.
- Plan to install and configure host multipath software to achieve host-side channel redundancy.
- Plan for management access to the storage platform by using either the in-band management or the out-of-band management methodology (out-of-band is most commonly used).

- Use the SANtricity ES client to connect to the storage system and to implement the planned configuration.
- Always save the system configuration and profile after configuration or provisioning changes so that in case of a catastrophic system fault the system can be fully recovered.

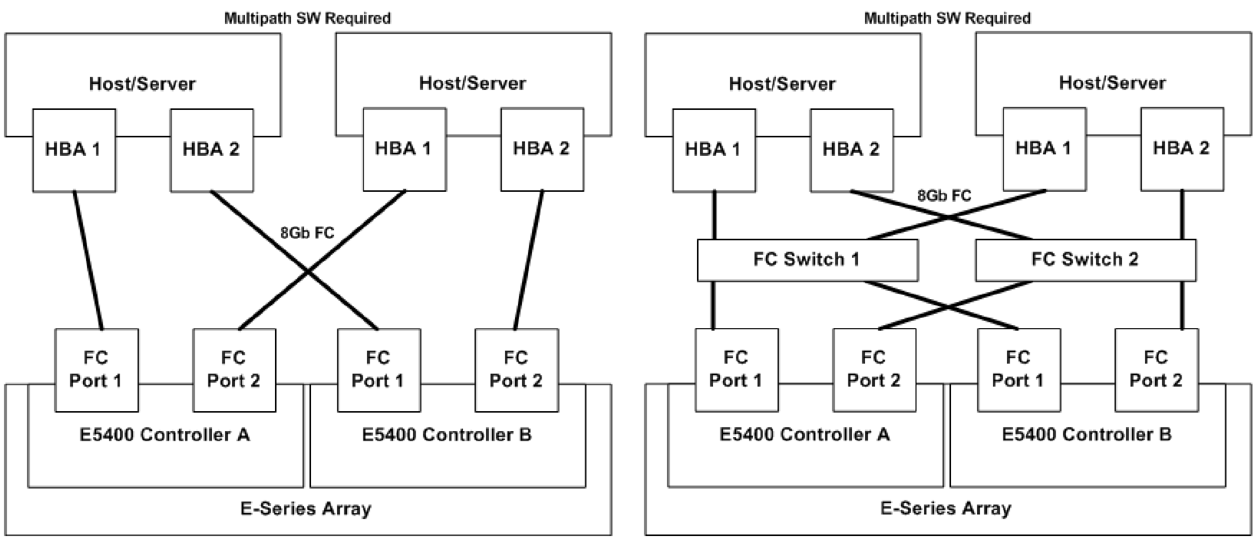
SANtricity ES is the GUI management interface for E-Series arrays. It is based on the Java framework and can be installed on Windows or Linux OSs. The management application should be installed on a management node that does not participate in production data delivery. The software is available in 32-bit and 64-bit versions, and the install process detects if the installation of the package is performed on the wrong OS version.

The SANtricity ES client software can be installed on Windows or Linux OS for out-of-band management of the storage array. In this configuration, the host agent functionality for in-band management does not function, and the number of client connections is limited to eight. To manage the storage arrays by using in-band connections, the management client must be running a server OS and have FC connectivity to all arrays. In this configuration, the eight-session maximum does not apply.

Additional Information

For host-side FC and IB connections, the hosts can be connected either directly to the storage controller or through a switch that allows multiple hosts to share the paths, as shown in Figure 11. Both configurations require multipath software on the host for link management.

Figure 11) Host connection examples.



4.2 E-Series E2600 Hardware

Overview

High-bandwidth applications and HPC platforms require high-performance, reliable, and scalable storage systems. The E2600-based storage system meets these requirements by supporting:

- Two 4-lane 6Gb SAS host interface ports per controller standard
- Multiple optional HICs, one per controller:
  - Four-port 8Gb FC
  - Four-port 1Gb iSCSI
  - Two-port 10Gb iSCSI

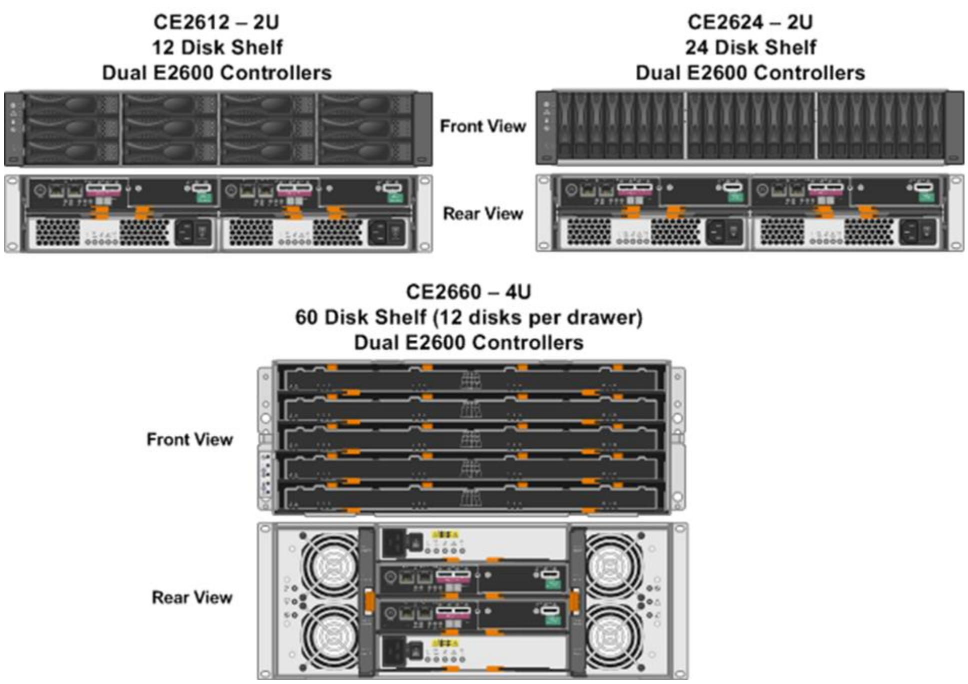
- Two-port 6Gb SAS

- 192 total disk drives per storage array
- Multiple RAID levels (0, 1, 10, 3, 5, and 6)
- A range of drive speeds and capacities
- Data assurance (T10-PI data integrity checking)
- Media parity check and correction capability
- Extensive event logging
- Recovery Guru onboard system diagnostics and recovery capability
- Hardware redundancy
- 1GB cache memory per controller (2GB optional) to maximize read/write performance
- NVSRAM and onboard USB drive to preserve the system configuration during power outages

As shown in Figure 12, the E2600 controller is supported in three shelf packages (E2660, E2624, and E2612), each supporting dual-controller canisters, power supplies, and fan units for hardware redundancy. The shelves are sized to support 60 disks, 24 disks, or 12 disks, respectively. Multiple disk expansion shelves (DE6600, DE5600, and DE1600) can be connected to the controller shelf to add additional storage capacity. For additional details, refer to the NetApp E2600 Storage System datasheet.

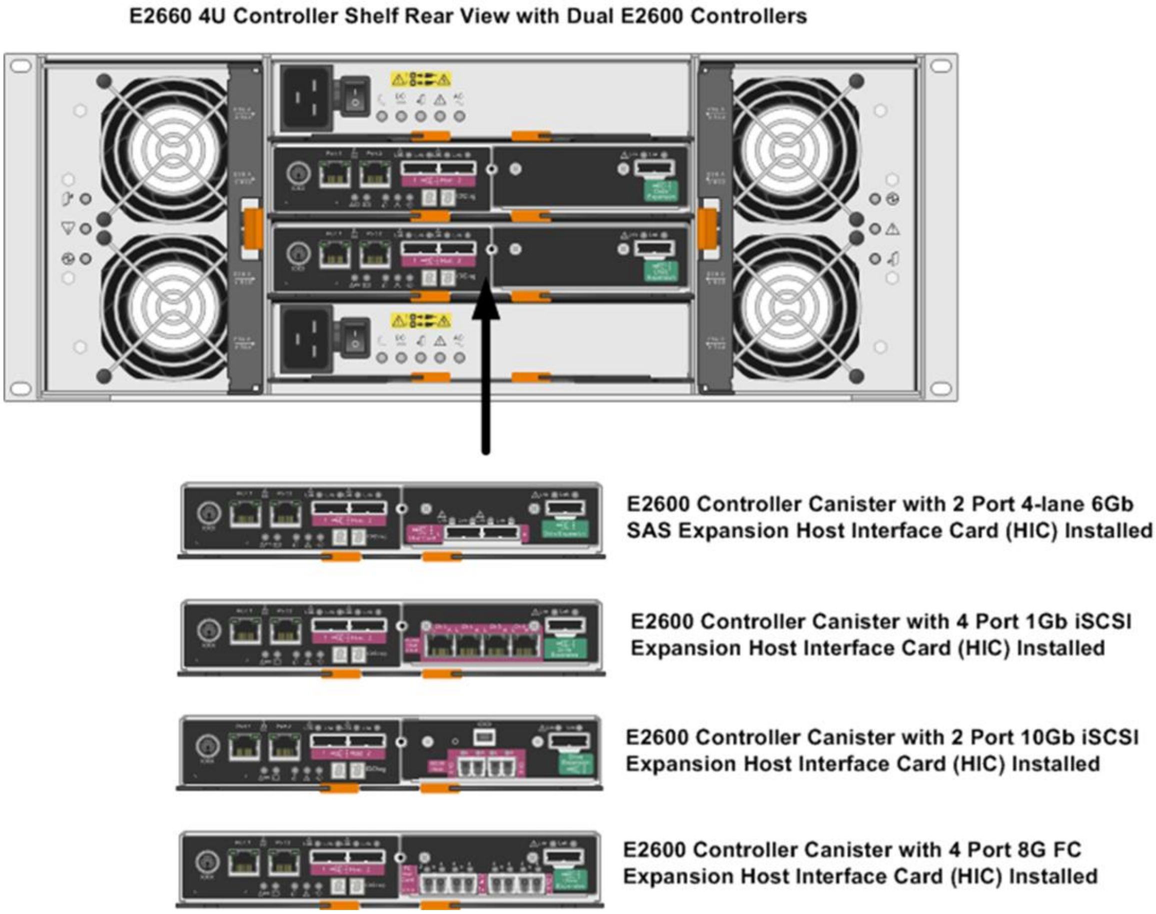
Note: Empty slots in any attached disk shelf are counted as drives when calculating the total drive count on an array.

Figure 12) E2600 shelf options.



By default, the E2600 controller canister has two onboard 4-lane 6Gb/s SAS ports for host-side communication channels, but it also supports host-side channel expansion through add-on modules that add either four FC ports and four 1Gb iSCSI ports, two 10Gb iSCSI ports, or two additional 4-lane 6Gb SAS ports. Figure 13 shows the E2660 4U shelf with the available controller configurations.

Figure 13) E2660 controller with optional host-side expansion ports.

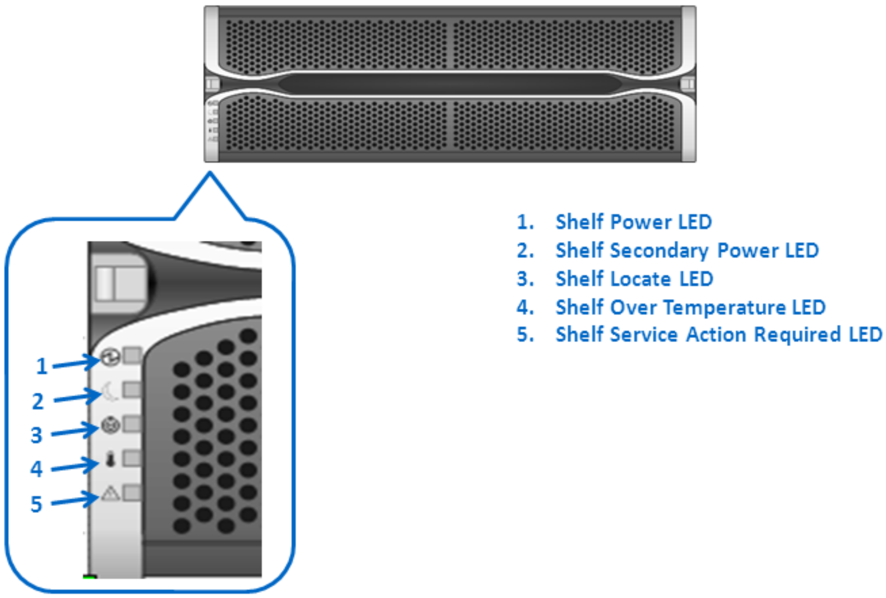


LED Status Indicators

Controller Drive Shelf LED Status Indicators

The E2600 controller shelf enclosure has several LEDs that indicate the overall status of the array, as shown in Figure 14.

Figure 14) Controller drive shelf status LEDs.



The shelf status LED layout is the same for all three packaging options (DE6600, DE5600, and DE1600). Table 18 lists the meanings of all indicators.

Table 18) Controller disk shelf LED status definitions.

LED Name	Color	LED On	LED Off
Controller Drive Shelf Power	Green	Power is present.	Normal status.
Controller Drive Shelf Secondary Power	Green	Battery is fully charged. LED blinks when battery is charging.	Controller canister is operating without battery, or existing battery has failed.
Controller Drive Shelf Locate	White	Identifies controller-drive tray when SANtricity ES Locate feature is activated.	Normal status.
Controller Drive Shelf Over Temperature	Amber	Temperature of controller-drive tray has reached unsafe level.	Normal status.
Controller Drive Shelf Service Action Required	Amber	Component within controller-drive tray requires attention.	Normal status.

Controller Base Features LED Status Indicators

The E2600 controller has several onboard LED status indicators, as shown in Figure 15. Most of the LEDs are lit when a fault condition exists; however, the battery-charging and the cache-active LEDs are lit when the battery is fully charged and the cache is active. The seven-segment LEDs provide status codes for both normal operation and fault conditions, and the dot in the first seven-segment LED is the controller heartbeat indicator.

Figure 15) E2600 controller status indicator LEDs.

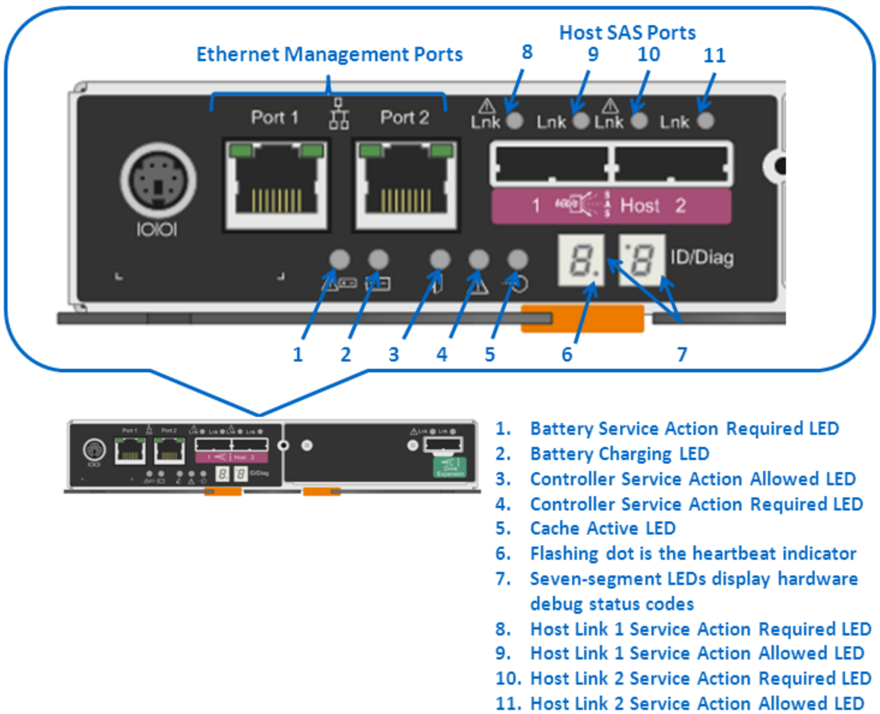


Table 19 provides additional controller status LED definitions.

Table 19) Controller base features LED status definitions.

LED Name	Color	LED On	LED Off
Battery Service Action Required LED	Amber	Battery in controller canister has failed.	Normal status.
Battery Charging LED	Green	Battery is fully charged. LED blinks when battery is charging.	Controller canister is operating without battery, or existing battery has failed.
Controller Service Action Allowed LED	Blue	Controller canister can be removed safely from controller-drive tray.	Controller canister cannot be removed safely from controller-drive tray.
Controller Service Action Required LED	Amber	Some fault exists within controller canister.	Normal status.
Cache Active LED	Green	Cache is active. After AC power failure, this LED blinks while cache offload is in process.	Cache is inactive, or controller canister has been removed from controller-drive tray.
Dot in Lower-Right Corner of First Seven-Segment LED	Yellow (not amber)	Flashing dot indicates controller heartbeat is active.	Dot not lit indicates controller heartbeat is not active (that is, controller is not in service).

LED Name	Color	LED On	LED Off
Two Seven-Segment LEDs	Yellow (not amber)	If controller status code = 99, then controller is in service. If controller status code does not = 99, then fault condition exists. Contact Technical Support for further assistance.	Controller is not powered on.

Note: The Battery Service Action Required LED indicates that the battery timer has expired or the battery has failed the automatic battery test. This condition can seriously impact the system write performance because the write cache feature is automatically disabled when the battery is not functioning normally.

Host-Side Ports LED Status Indicators

The host-side connection ports provide status LEDs to indicate the connection status for each link between the storage array and various host-side hardware devices, as shown in Figure 15. Table 20 and Table 21 provide the definition for each LED.

Table 20) Ethernet management port status indicator definitions.

LED Name	Color	LED On	LED Off
Ethernet Management Port Link Rate LED (top left corner of Mgmt. port RJ-45 connectors)	Green	There is a 100BASE-T rate.	There is a 10BASE-T rate.
Ethernet Management Port Connectors Link Active LED (top right corner of Mgmt. port RJ-45 connectors)	Green	Link is up (LED blinks when there is activity).	Link is not active.

Table 21) Host-side SAS ports status indicator definitions.

LED Name	Color	LED On	LED Off
Host Link 1 Service Action Required LED	Amber	At least one of four PHYs is working, but another PHY cannot establish same link to device connected to Host IN port connector.	No link error has occurred.
Host Link 1 Service Action Allowed LED	Green	At least one of four PHYs in Host IN port is working, and link exists to device connected to IN port connector.	No link error has occurred.
Host Link 2 Service Action Required LED	Amber	At least one of four PHYs is working, but another PHY cannot establish same link to device connected to Host IN port connector.	No link error has occurred.

LED Name	Color	LED On	LED Off
Host Link 2 Service Action Allowed LED	Green	At least one of four PHYs in Host IN port is working, and link exists to device connected to IN port connector.	No link error has occurred.

Drive-Side SAS Expansion Port

The E2600 controller canister is equipped with a SAS expansion port used to connect additional disk shelves to the E2600 controller shelf. Figure 16 shows a close-up of the SAS expansion port LEDs.

Figure 16) E2600 drive expansion port status indicator LEDs.

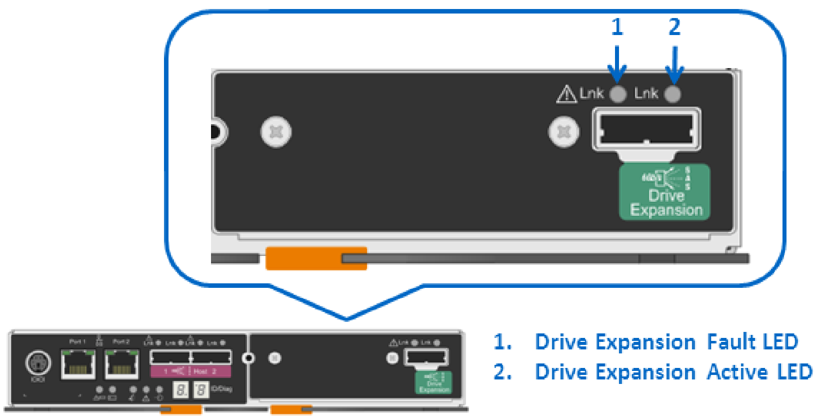


Table 22 provides the definition for each drive-side LED.

Table 22) Drive-side SAS ports status indicator definitions.

LED Name	Color	LED On	LED Off
Drive Expansion Link Fault	Amber	At least one of four PHYs in Out port is working, but another PHY cannot establish same link to Expansion Out connector.	Normal status
Drive Expansion Link Active	Green	At least one of four PHYs in Out port is working, and link exists to device connected to Expansion Out connector.	Link error has occurred

For additional details on the E2600 controller and related hardware, refer to the NetApp E-Series Storage Systems CE2600-60 Controller-Drive Tray Installation Guide.

Guidelines

Consider the following guidelines when implementing the E2600 storage system:

- Determine the level of performance required by the compute platforms to support the given applications.

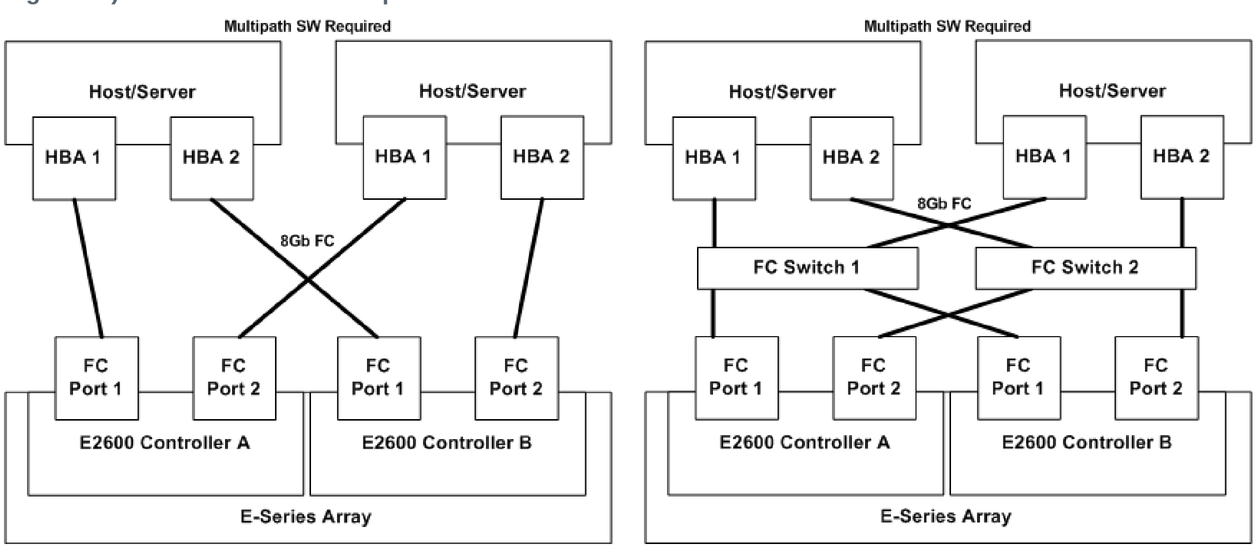
- Determine the amount of storage capacity required (include the number of disks required for hot spares).
- Choose the disk types based on performance and capacity requirements.
- Determine the power and network connectivity requirements.
- Plan RAID levels to achieve the level of reliability and read/write performance required.
- Determine which hosts will be connected to the storage system, and plan the configuration of the storage system ports to maximize throughput.
- Plan to install and configure host multipath software to achieve host-side channel redundancy.
- Plan for management access to the storage platform by using either the in-band management or the out-of-band management methodology (out-of-band is most commonly used).
- Use the SANtricity ES client to connect to the storage system and to implement the planned configuration.
- Always save the system configuration and profile after configuration or provisioning changes so that in case of a catastrophic system fault the system can be fully recovered.

SANtricity ES is the GUI management interface for E-Series arrays and operates in an in-band or in an out-of-band mode, but the management application should be installed on a management node that does not participate in production data delivery. The GUI is based on the Java framework and can be installed on Windows or Linux OSs. The software is available in 32-bit and 64-bit versions. The install process detects the OS and hardware version and installs the correct version for that platform. To manage the storage arrays by using in-band connections, the management client must be running a server OS and have FC connectivity to all arrays.

Additional Information

For host-side FC and iSCSI connections, the hosts can be connected either directly to the storage controller or through a switch that allows multiple hosts to share the paths, as shown in Figure 17. Both configurations require multipath software for link management.

Figure 17) Host connection examples.



SAS ports are usually cabled directly to local servers. In this configuration, determine that all host servers have a path to both controllers in E-Series arrays, and install the appropriate multipath software for the server OS type.

4.3 E-Series Disk Expansion Shelves

Overview

E-Series arrays support storage capacity growth beyond the disk slots in the controller shelf by adding disk expansion shelves to new or existing E5400- and E2600-based storage arrays. The additional DE6600 (60-disk), DE5600 (24-disk), or DE1600 (12-disk) shelf enclosures have environmental services monitor (ESM) canisters installed instead of controller canisters. Figure 18 shows the ESM canister.

Figure 18) ESM canister.

E-Series disk expansion shelves can be added in combinations of 4U and 2U packages to achieve specific performance and capacity requirements. The typical configurations for each shelf type shown in Figure 19, Figure 20, and Figure 21 represent best-practice cabling topology to maximize system resiliency against shelf hardware fault scenarios.

Note: The ESM canister has two SAS input ports and one output port for intershelf cabling. Never connect to both input ports on an ESM canister. Only one of the input ports can be used.

Figure 19) Maximum capacity E-Series array configuration using DE6600 shelves.

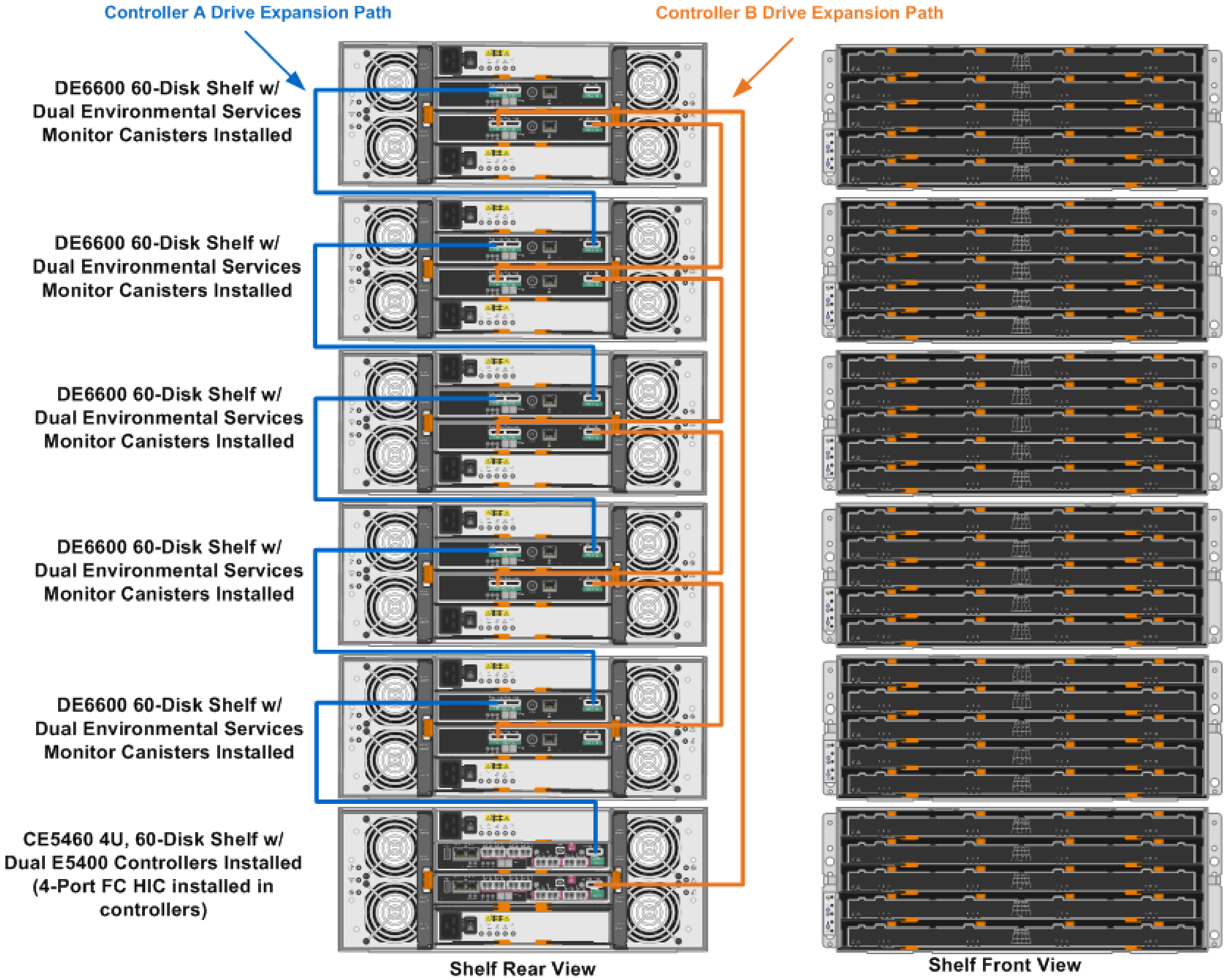


Figure 20) Typical E-Series array configuration using DE5600 shelves.

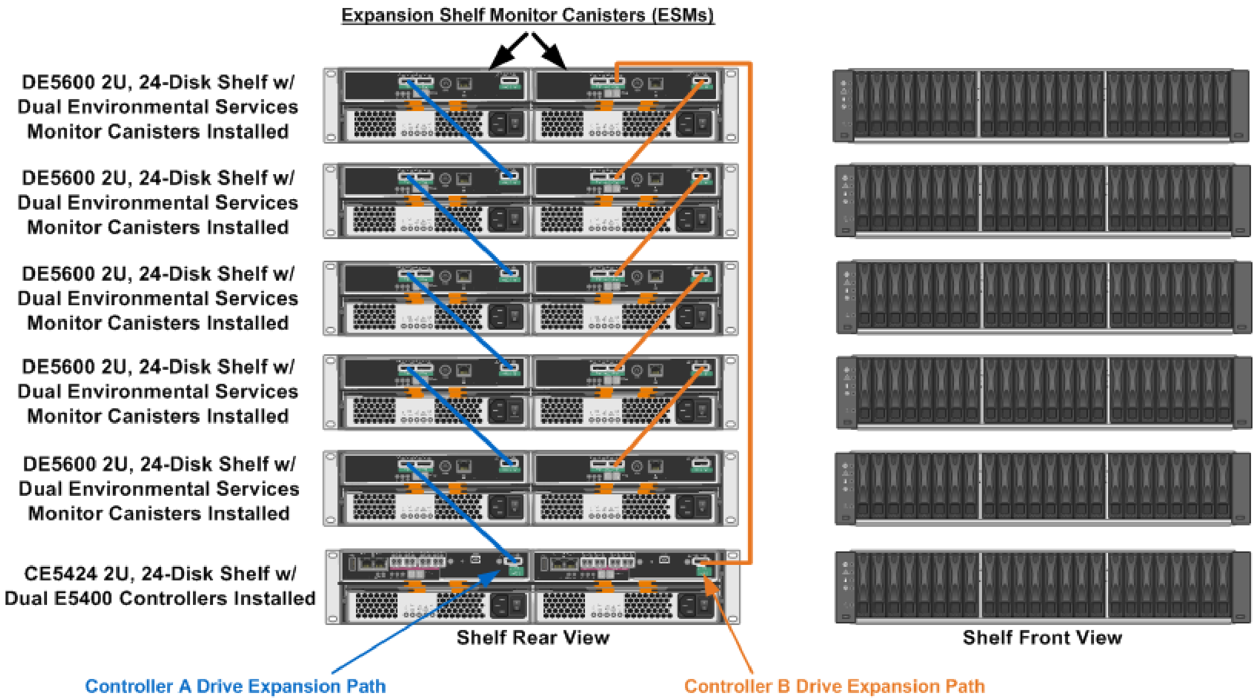
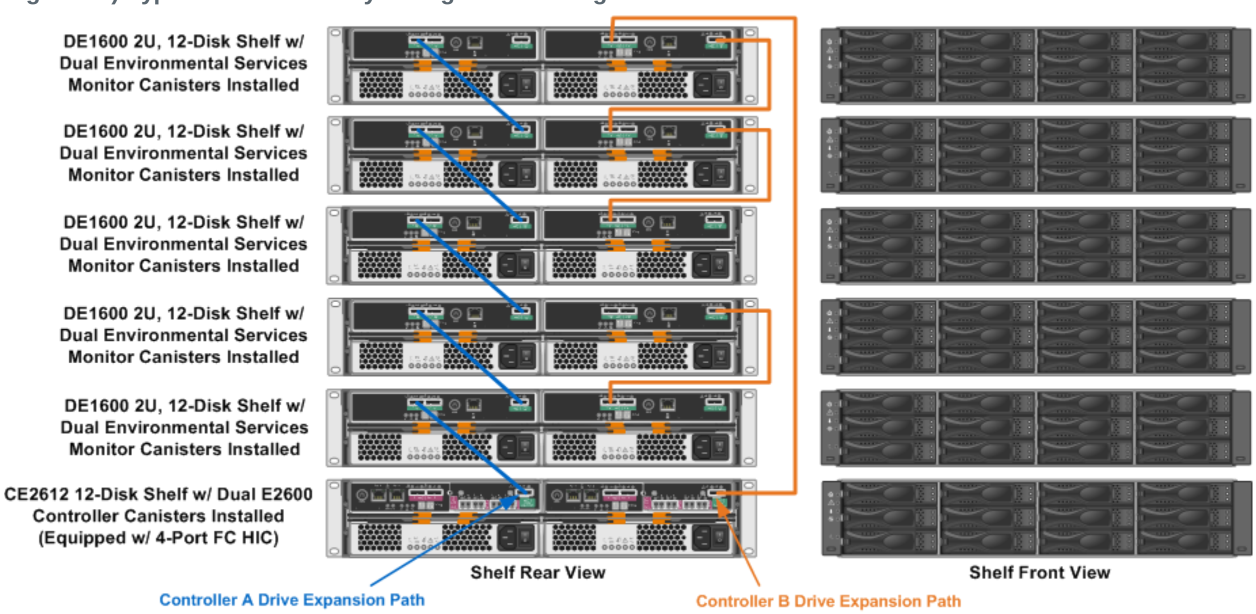


Figure 21) Typical E-Series array configuration using DE1600 shelves.



Note: NetApp does not recommend mixing shelf models in the same array because of the differences in the drives and the supported features.

The disk shelves must be installed within six feet of the array controller shelf to allow SAS cables to reach from the SAS expansion ports on the controller canisters to the ESM canisters installed in a disk shelf, or from the SAS expansion port on one expansion shelf to the ESM SAS input port on a successive disk expansion shelf.

For additional details on how to install the E5400-based array configurations, refer to the NetApp E-Series Storage Systems CE5400 Controller-Drive Tray Installation Guide. For additional details on how to install the E2600-based storage array, refer to the NetApp E-Series Storage Systems CE2600-60 Controller-Drive Tray Installation Guide.

The most common configuration for an array is to stack the disk expansion shelves in the same physical rack that contains the controller shelf. However, care should be taken to ascertain that floor loading limitations are not exceeded. Whenever possible, the 60-disk chassis should be placed in the lower portion of the racks. Appropriate lifting equipment should be used to mount the chassis because a single shelf with the disks installed can exceed 200 pounds.

The E-Series E5400-based storage array can support a maximum of 384 individual drives while the E2600-based storage array can support up to 192 individual drives. However, depending on shelf model and disk selection, additional boundaries to drive count must be considered. For more information on supported configurations, refer to the NetApp E5400 Storage System datasheet or to the NetApp E2600 Storage System datasheet.

Note: Empty disk slots in any connected disk shelf still count as a disk for planning the total disk count for a single storage array.

When initially powering on an E-Series array with disk expansion shelves, power on the disk shelves first and wait one to two minutes before powering on the controller shelf. To add a disk expansion shelf to an existing E-Series array, follow the specific installation steps. For more information and assistance with adding a disk expansion shelf to an existing production E-Series array, contact NetApp Global Services.

Guidelines

Follow these best practices when installing and configuring the E5400 storage system:

- Do not mix different drive speeds in the same drive shelf drawer.
- Do not use both SAS input ports on an ESM at the same time.
- Install the CE5460 shelves at the bottom of the racks to prevent the rack from becoming top heavy.
- Use lifting equipment when mounting controller and disk shelves because these shelves can exceed 200 pounds when they are fully loaded with disks.
- Replace failed disks with disks that match the failed disk:
  - SSDs must be replaced by other SSDs.
  - Encryption-capable drives must be replaced by other encryption-capable drives.
  - Drives supporting T10PI data assurance must be replaced by T10PI-capable drives.
- Route disk channel cables to avoid single points of failure.

Note: For additional cabling guidance for specific storage array models, refer to the NetApp E-Series Storage Systems Hardware Cabling Guide on the NetApp Support site.

5 Storage for E-Series

5.1 E-Series Quantum StorNext SAN Client Storage Configuration

Overview

NetApp E-Series solutions that use SNFS share the E-Series storage between the SAN client nodes such that the FC storage appears to be locally attached. Metadata and file system journal information is passed over an Ethernet connection, but the actual data is written directly from the SAN client to the storage. Use SANtricity ES to create six RAID 6 (8+2) volume groups per 60-disk array. Use the entire capacity of the volume group to create one volume in each group, and then map each LUN to a host group containing all of the SAN client nodes.

Note: RAID 6 provides protection from dual-disk failure scenarios. However, when all available disks in the array are allocated in volume groups, NetApp recommends treating any disk failure scenario as a critical service event. Therefore, purchase and store spare disks for easy access, and replace faulty disks as soon as possible after failure.

A volume group should contain disks of the same type and capacity. For best performance, keep all cache settings at their default. When configuring the array, use a standard naming convention that associates the volume group and the volume names to the unit and LUN ID. This convention is helpful when troubleshooting E-Series solution path faults.

Before configuring the array, verify that no default volume groups or volumes are preconfigured on the array. Remove existing volumes or volume groups, and make sure that all 60 disks on the CE6600 and all 24 disks on the CE5600 disk shelves are unassigned. This verification is required because the base E-Series configuration for use with SNFSs requires the entire capacity of the array for client data.

For additional information about file system planning, refer to the “Quantum StorNext Installation and Maintenance Guide” located on the Quantum Web site.

Guidelines

Follow these guidelines to determine the appropriate configuration for E-Series solutions that use the SNFS:

- Determine if the disk initialization time for volume group creation is a concern. Although the Data Assurance feature is available for certain types of disks and is extremely useful in some environments, it greatly increases the time needed for disk initialization.
- Use SANtricity ES and the Automatic Configuration feature to create six RAID 6 (8+2) volume groups on the CE6600 shelf and three RAID 5 (7+1) volume groups on the CE5600, with one volume per group consisting of the entire capacity of the group.
- When using the CE5600 shelf in an environment in which availability and tolerance to drive failures are the primary concerns, optionally use three 6+2 RAID 6 volume groups.
- Name the volumes such that the mapped LUN ID and preferred controller path are obvious from the naming convention for easier troubleshooting.
- Always start mapping clients to storage by using LUN 1 on the first array in the SNFS, and sequentially increase LUN numbers as the storage infrastructure for the file system increases (that is, only have one LUN 1).
- Monitor the system by using SANtricity ES, and set up alerts so that fault conditions are detected and addressed in a timely manner.
- Purchase and store spare disks so they are readily available to replace a faulty disk.
- Since the configuration of six RAID 6 (8+2) volume groups on the CE6600 shelf or three RAID 5 (7+1) or RAID 6 (6+2) volume groups on the CE5600 shelf does not allow room for hot spare disks on the array, treat any disk-related fault condition as a critical service event and replace the faulty disk as soon as possible.

5.2 E-Series Quantum StorNext Metadata Storage Configuration

Overview

When the NetApp E-Series E2624 array is used to satisfy the metadata and journal storage requirement for an SNFS, NetApp recommends that disks be mirrored together in a 2+2 RAID 10 set. One metadata and journal stripe group set is required for each file system. To meet the MDC storage requirement for a stripe group, use one RAID 10 volume group consisting of four disks. The logical unit number (LUN) volume must be constructed using the entire capacity of the volume group.

Note: RAID 10 provides protection from disk failure scenarios by writing a duplicate copy of the data on each disk to a second disk, thus forming mirrored pairs of disks. If all available disks in the array are allocated in volume groups, NetApp recommends that any disk failure scenario be treated as a critical service event and that faulty disks be replaced as soon as possible after failure. Given this requirement, customers should purchase and store spare disks in case a disk-related service event occurs.

Use the SANtricity ES Automatic Configuration feature to provision the volume group and volume, and map the volumes to either the default host group or to a specific host group.

Guidelines

Follow these guidelines when determining the appropriate MDC configuration for E-Series solutions that use the SNFS:

- Confirm that each MDC is associated either with the default group or with the appropriate host group on the E-Series array. The metadata/journal stripe group LUNs do not need to be accessible through FC by the StorNext clients.
- Use the Automatic Configuration feature to configure the volume group and volume.
- Use a segment size of 128K (typical file system setting) and accept the default caching options.
- Monitor the system by using SANtricity ES and set up alerts so that fault conditions are detected and addressed in a timely manner.
- Allocate at least one spare disk per 24-disk array, or purchase spare disks to have them readily available in case a disk failure occurs.
- This configuration uses RAID 10 disk protection. Therefore, the volume group is protected from multiple disk failure scenarios; however, for failure scenarios in which both disks in any mirrored pair fail, the file system data is lost. NetApp strongly recommends treating any disk failure scenario as a critical service event.

Note: When the E2624 is used to support more than one SNFS, (that is, more than one metadata/journal LUN is provisioned on the E2624), alternate RAID controller LUN ownership between the LUNs to maximize system performance.

6 Operating Systems Connecting to E-Series

6.1 E-Series Quantum StorNext SAN Client Host Configuration

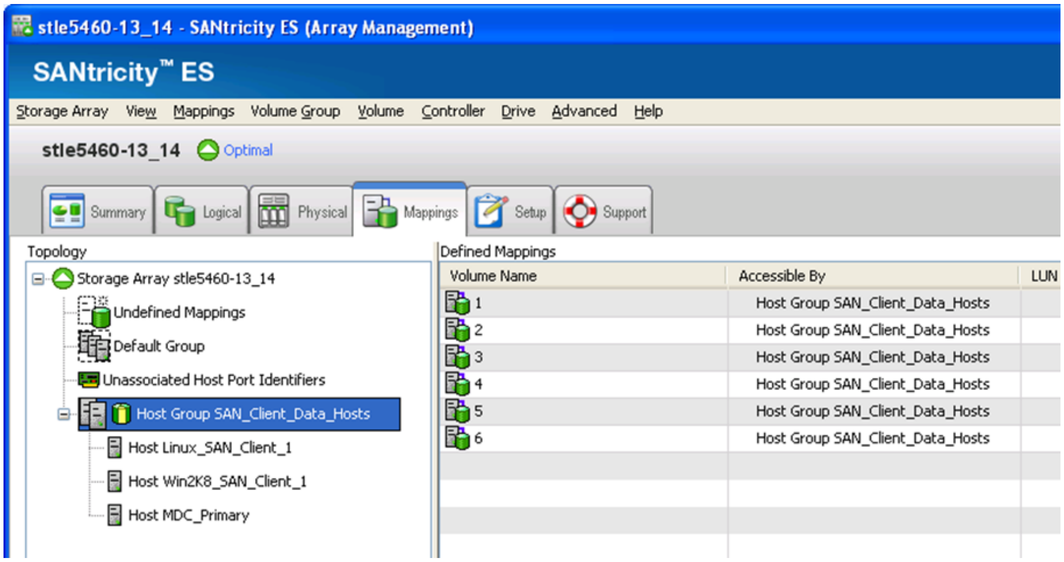
Overview

NetApp E-Series solutions that use the SNFS require the SAN clients to be connected through an Ethernet connection (preferably a private switched network) to the MDC for metadata traffic and through FC connections to the E-Series storage for direct block data transfers. The SAN client host configuration consists of two parts. The first is the configuration on the E-Series storage arrays, and the second is the configuration of the host to mount the SNFS.

The first part of the configuration uses the standard E-Series host creation procedure. In this procedure, each path to a host is identified by the associated World Wide Port Name (WWPN) and labeled in the array with an alias that aids the manageability of the paths over time. When prompted during the creation procedure, indicate that the SAN client host will share storage with other hosts. This allows the SAN client host to be associated with a host group that is mapped to all the available storage capacity on the array.

When creating the first SAN client host on the E-Series array, enter a new host group name to begin the host associations. This configuration enables the full support for mixed operating system client environments and represents one of NetApp’s best practices for E-Series storage access security. To create additional SAN client hosts on the array, select the host group with the first host and define the new host as shown in Figure 22. This method adds the new host to the desired host group. The group is then mapped to all of the available LUNs on the array to form a storage partition configuration that is indicated by the yellow cylinder in the host group icon.

Figure 22) E-Series host group configuration example.



In environments with multiple E-Series arrays in the same file system, the first array can be manually configured. The associated configuration file can be used to script the setup of other E-Series arrays in the file system.

To enable the SAN clients to mount the SNFS, each node must have the appropriate StorNext software installed and be connected to both the Ethernet network for metadata traffic and to the FC fabric for block data transfer to the E-Series array. When a client has multiple FC links to the array, install the appropriate multipath driver to manage the multiple paths. The following multipath drivers are commonly used with StorNext clients:

- Linux. Use SANtricity ES RDAC or Linux DMMP with NetApp devices enabled.
- Windows. Use NetApp SANtricity MPIO-DSM.
- Solaris. Use Sun-MPxIO.
- ESX®. Use VMware®-specific failover built in to ESX.
- Other multipath drivers require proof-of-concept testing to verify specific version compatibility.

Depending on each SAN client's needs, Windows server, and hardware configuration, different mount options may be used to optimize performance for that node and for the associated workflows. For most Linux systems with more than 8GB of RAM and 8 CPUs, the following mount options may be used:

- `rw`
- `cachebufsize=1024k` (for RAID 6 [8+2])
- `buffercachecap=8192`
- `buffercache_readahead=64`
- `threads=16`
- `sparse=yes`

Note: `cachebufsize` is derived from multiplying the number of data disks in a volume group by the stripe size, which is 128K when using the default recommended LUN stripe size for most E-Series and StorNext solutions. As a result, the `cachebufsize` for a RAID 5 (7+1) LUN assuming a 128K stripe size is 7×128K=896K, and a 6+2 RAID 6 LUN assuming a 128K stripe size is 6×128K=768K.

For Linux servers, these options are specified in the `/etc/fstab` file on each SAN client server, and they do not have to be consistent between the servers mounting the file system.

Guidelines

NetApp recommends the following guidelines:

- Carefully plan the metadata LAN connectivity and FC fabric to avoid single points of failure.
- Use the SANtricity ES Host Creation wizard to define a new host group with the first host and associate additional hosts to that host group.
- Use the appropriate multipath drivers when there are multiple links to an E-Series array.
- Use SANtricity ES to provide alerts of any hardware failures or issues as soon as possible.

6.2 E-Series Quantum StorNext Metadata Host Configuration

Overview

A StorNext metadata environment may be constructed by using the Quantum M330 appliance as the MDC, or the customers may provide their own MDC servers and use an E-Series E2624 as the storage device for the metadata and journal stripe groups.

When the E-Series E2624 is used, it is usually SAS-attached directly to the MDC host through a SAS HBA. Optionally, it can be FC-attached to the MDC host if the E2600 controller is equipped with an FC expansion HIC. When an MDC server has multiple connections to the array, install the appropriate multipath driver for the specific operating system.

The MDC hardware must meet the requirements listed in the Quantum StorNext installation documentation located on the Quantum Web site, and the host interfaces must meet the requirements for E-Series products listed in the NetApp IMT.

Like the StorNext SAN client, the MDC host configuration consists of two parts. The first is the configuration on the E-Series storage arrays, and the second is the configuration of the SNFS software on the server.

The first part of the configuration uses the standard E-Series host creation procedure. In this procedure, each path to a host is identified by the associated World Wide Port Number (WWPN) and labeled in the array with an alias that aids the manageability of the paths over time. The procedure requires administrators to access the MDCs, retrieve the WWPNs, and then match the mapping to the ports with the specified WWPNs.

When creating the metadata host on the E-Series array, indicate that the metadata host will share storage with other hosts. This allows the metadata host to be associated with a storage partition containing all of the available storage capacity of the metadata and journal volume group (in this case, the single metadata and journal LUN) and also the secondary MDC when using a high-availability configuration. Create the host group as part of the procedure for creating the MDC host on the array. This configuration enables the E2624 to support multiple SNFSs and represents one of NetApp’s best practices for E-Series storage access security.

In addition to the MDC path to the metadata and journal LUN, the MDC also requires access to the client data storage LUNs for management access. This is accomplished by installing at least one physical FC link from the MDC that is zoned with one of the array host interfaces for each array in the file system. On each E-Series client data storage array in the file system, create a host for the MDC and map the host to the host group with the other StorNext SAN clients. This connection is primarily used to label the disks (LUNs) as one of the steps to create the SNFS, but it should never be used for metadata storage workflows. When creating the file system on the MDC, be careful not to select this storage as the metadata and journal LUN. The metadata and journal LUN should only be located on the E2624 array.

The second part of the MDC host configuration is addressed by the "Quantum StorNext Installation and Implementation Guide" in the Quantum Web site.

Guidelines

Follow these guidelines when determining the appropriate MDC storage configurations for E-Series solutions that use the SNFS:

- Install and configure MDC servers by using the standard practices for those devices.
- Create the MDC on all E-Series arrays in the file system by using the SANtricity ES management application procedure to create hosts.
- Install the StorNext software on the MDCs by using the standard Quantum installation procedure.
- On the client data storage arrays, add the MDC host to the SAN client host group as part of the host creation procedure.
- On the E2624 array with the metadata and journal LUN, create a new host group as part of the host creation procedure.
- On the E2624, map the new host group to the metadata and journal and use the LUN 0 designation.
- For Linux MDCs, use the `/opt/SMgr/util/SMdevices` command to determine the SSN, WWN, and device ID for the newly added LUNs.
- Install the appropriate multipath drivers for the specific MDC servers.
- NetApp recommends maintaining the FC connectivity from the MDC to the client data storage arrays for future maintenance; however, this connection should not be used for metadata storage workflows.
- Take care to identify the correct metadata and journal LUN when configuring the file system on the MDC.

6.3 E-Series Quantum StorNext File System Installation

Overview

The Quantum StorNext application is commonly used for high-performance shared workflow operations and multitier archives. It has two components:

- StorNext File System (SNFS)
- StorNext Storage Manager (SNSM)

This document focuses only on SNFS. For more information on SNSM, refer to the product documentation on the Quantum Web site.

Quantum SNFS allows applications to work from a single consolidated dataset, which helps to streamline processes and facilitates fast job completion. SNFS applications that run on different OSs can simultaneously access and modify files in a common high-speed storage pool.

Quantum StorNext Installation

This section provides guidelines for installing Quantum's SNFS on a Linux MDC. If the Storage Manager components and the StorNext Linux GUI are not needed, perform an RPM Package Manager (RPM)–only installation. The Storage Manager component of Quantum StorNext can be installed only on Linux systems, but it is not within the scope of this document. For additional information about RPM-only installation or about installing Quantum SNFS on Windows, refer to the Quantum StorNext Installation Guide.

After Quantum SNFS has been successfully installed, it can be configured by using the Quantum StorNext Configuration wizard. Configuring Quantum StorNext consists of entering license information and creating one or more file systems. To access the Configuration wizard, use the URL for the Quantum StorNext GUI running on the MDC.

Before installing Quantum StorNext software, perform the following tasks:

- Verify that the MDC meets all OS and hardware requirements. For more information, refer to the Quantum StorNext Installation Guide.
- Verify that the storage devices are correctly configured and visible to the MDC.
- Download the file system software to the MDC server.
- Optional: Run the Quantum StorNext preinstallation script to check for available disk space and view recommended locations for support directories.

Quantum StorNext Configuration Using StorNext GUI

When the Quantum StorNext GUI is first opened, it launches the Quantum StorNext Configuration wizard. This wizard provides step-by-step guidance to configure Quantum StorNext. The Configuration wizard is optional. It consists of nine tasks and it tracks progress as each task is completed. The tasks are labeled on tabs in the wizard navigation pane:

- Welcome shows the disks and libraries currently available for Quantum StorNext usage.
- Licenses is used to enter Quantum SNFS license information.
- Name Servers is for specifying and ordering the machines acting as Quantum StorNext SAN client servers (name servers).
- File Systems can be used to add a Quantum SNFS.
- Storage Destinations can be used to add a library, storage disk, or other storage destination.
- Storage Policies is for adding a Storage Manager or replication storage policy.
- Email Server is used to specify an e-mail server for handling Quantum StorNext notifications.
- Email Notifications is used to add e-mail notification recipients.
- Done indicates that the Configuration wizard is complete. At completion, the installation can also be converted to a high-availability system.

File System Block Size, Metadata Disk Size, and Journal Size Settings

Advanced parameters that must be configured during the creation of a file system include `FsBlockSize`, metadata disk size, and journal size settings. These settings all work together. For example, the `FsBlockSize` setting must be correct in order for the metadata disk size and journal size to be correct.

FsBlockSize and Metadata Disk Size Setting

Optimal `FsBlockSize` settings for both performance and space utilization range from 16K to 64K. Avoid values greater than 64K; they can adversely impact performance because of inefficient metadata I/O operations. Avoid values of less than 16K because they can adversely impact start-up and failover time. Set the `FsBlockSize` to a higher value because it is important for multiterabyte file systems to obtain optimal start-up and failover times.

The minimum value for metadata disk size is 25GB. More space can be allocated, depending on the number of files per directory and on the size of the file system. Table 23 lists the suggested `FsBlockSize` and metadata disk space settings, which are based on the average number of files per directory and on file system size. The amount of metadata disk space is in addition to the minimum 25GB. Use Table 23 to determine proper configuration settings.

Table 23) Recommended FsBlockSize and metadata disk space settings.

Average Number of Files per Directory	File System Size: Less Than 10TB	File System Size: 10TB or Larger
Less than 10	FsBlockSize: 16KB Metadata: 32GB per 1M files	FsBlockSize: 64KB Metadata: 128GB per 1M files
10–100	FsBlockSize: 16KB Metadata: 8GB per 1M files	FsBlockSize: 64KB Metadata: 32GB per 1M files
100–1,000	FsBlockSize: 64KB Metadata: 8GB per 1M files	FsBlockSize: 64KB Metadata: 8GB per 1M files
More than 1,000	FsBlockSize: 64KB Metadata: 4GB per 1M files	FsBlockSize: 64KB Metadata: 4GB per 1M files

Best Practice
Use a 64K `FsBlockSize` unless requirements dictate otherwise. The `FsBlockSize` setting is not adjustable after the file system is created. Therefore, it is important to give careful consideration to this setting during the initial configuration.

Journal Size Setting

The optimal journal size setting ranges from 16MB to 64MB, depending on the `FsBlockSize` value. Avoid values larger than 64MB because they can cause potentially severe impacts on start-up and failover times. Values on the higher end of the 16MB to 64MB range can improve performance of metadata operations in some cases but can cause slower start-up and failover time. The journal size for new file systems should be at least 1,024 times the `FsBlockSize` size.

Initial File System Configuration Settings

Quantum SNFSs are highly tunable to meet varying workloads. To maximize performance on E-Series arrays, use the Quantum StorNext GUI to set up the file system parameters, as shown in Table 24.

Table 24) Quantum SNFS initial configuration settings.

Location	Variable	Recommended Value	Notes
Advanced Parameters > Allocation tab	File System Block Size	64KB	
	Journal Size	128MB	See journal sizing guide
	Strategy	Round Robin	Default
	Reserved Space	Checked	Default
	Stripe Align Size	-1	Default
	Inode Stripe Width	0	Default
	Allocation Session Reservation	Not checked	Default

Location	Variable	Recommended Value	Notes
Advanced Parameters > Performance tab	Max Connections	32	This value should be set to at least 2x the number of clients mounting the file system
	Buffer Cache Size	512MB	
	Inode Cache Size	512KB	
	Thread Pool Size	512	
	Use Physical Memory Only	Not checked	Default
	High-Priority FSM	Not checked	Default
Stripe Group/Disk Management > Metadata and Journal Volume	Name	`sg0`	The metadata and journal stripe group are typically `sg0`
	Breadth	64k	This value should be the same as the file system block size used
	Metadata	Checked	
	Journal	Checked	
Stripe Groups > User Data Volumes	Name	`sg1`, `sg2`, `sg3`, and so forth	
	Breadth	1MB	This is very important to set
	User Data	Checked	

Note: NetApp recommends enabling remote log-in access to the MDCs during the implementation phase of deployments to facilitate the file system setup activities.

RAID Stripe Size and Stripe Breadth

Configuration settings, such as RAID level, segment size, stripe size, and stripe breadth, are important and cannot be changed after the file system is put into production. Therefore, it is critical to determine appropriate settings during the initial configuration.

The StorNext stripe breadth size for the user data stripe groups is determined by multiplying the number of data disks in the RAID group by the segment size. For example, an 8 + 2 RAID 6 volume group with a 128KB segment size results in a 1MB stripe breadth size. The stripe breadth size is a critical factor for write performance because I/Os smaller than the stripe size can incur a read-modify-write penalty.

For additional information about these and other settings, refer to the [Quantum StorNext File System Tuning Guide].

Guidelines

When planning the implementation of a Quantum SNFS, consider the following guidelines:

- Verify that the MDC can communicate with:
  - The StorNext client nodes, using an isolated local IP network
  - The metadata stripe group

- Complete the file system planning and verify that all application requirements are accounted for.
- Verify that the server and storage settings are correct. This is important because many infrastructure settings cannot be changed once the file system is started.
- Before starting the file system:
  - Label the data disks properly.
  - Complete all of the file system initial configuration settings using the Quantum StorNext GUI.

6.4 E-Series Quantum StorNext Client Software Installation

Client servers can mount the SNFS by installing an OS-specific software package and completing the steps required by the OS to mount the file system. This section provides information on the high-level steps needed to acquire the correct software for a Quantum StorNext client-only installation on all supported OS types. The instructions to mount and work with a Quantum SNFS using the Quantum StorNext client software are specific to the OS type. This section covers only the description of how to put the SNFS and client software on a StorNext SAN client.

To run the Quantum StorNext client software, the client system must meet all OS and hardware requirements. For more information, refer to the Quantum StorNext Installation Guide.

Quantum StorNext clients are available for the following OSs:

- Apple® Mac® OS X
- HP-UX
- IBM AIX
- Red Hat Enterprise Linux 4
- Red Hat Enterprise Linux 5
- SUSE Linux Enterprise Server 10
- SUSE Linux Enterprise Server 11
- Sun Solaris 10
- Windows 7
- Windows Server® 2003
- Windows Server 2008
- Windows Vista®
- Windows XP

SAN clients must meet the following minimum hardware requirements:

- 1GB RAM
- 500MB available hard disk space

Acquire Quantum StorNext Client Software

The Quantum StorNext client software is available for download from:

- The Quantum StorNext installation DVD
- The Internet
- An MDC running the Quantum StorNext GUI

Download client software from an MDC only if the Quantum SNFS and SNSM are installed and running the Quantum StorNext GUI. Do not download client software from a Red Hat 4 or Windows MDC, or from a file-system-only Red Hat 5 MDC.

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

Go further, faster®

NetApp®

www.netapp.com